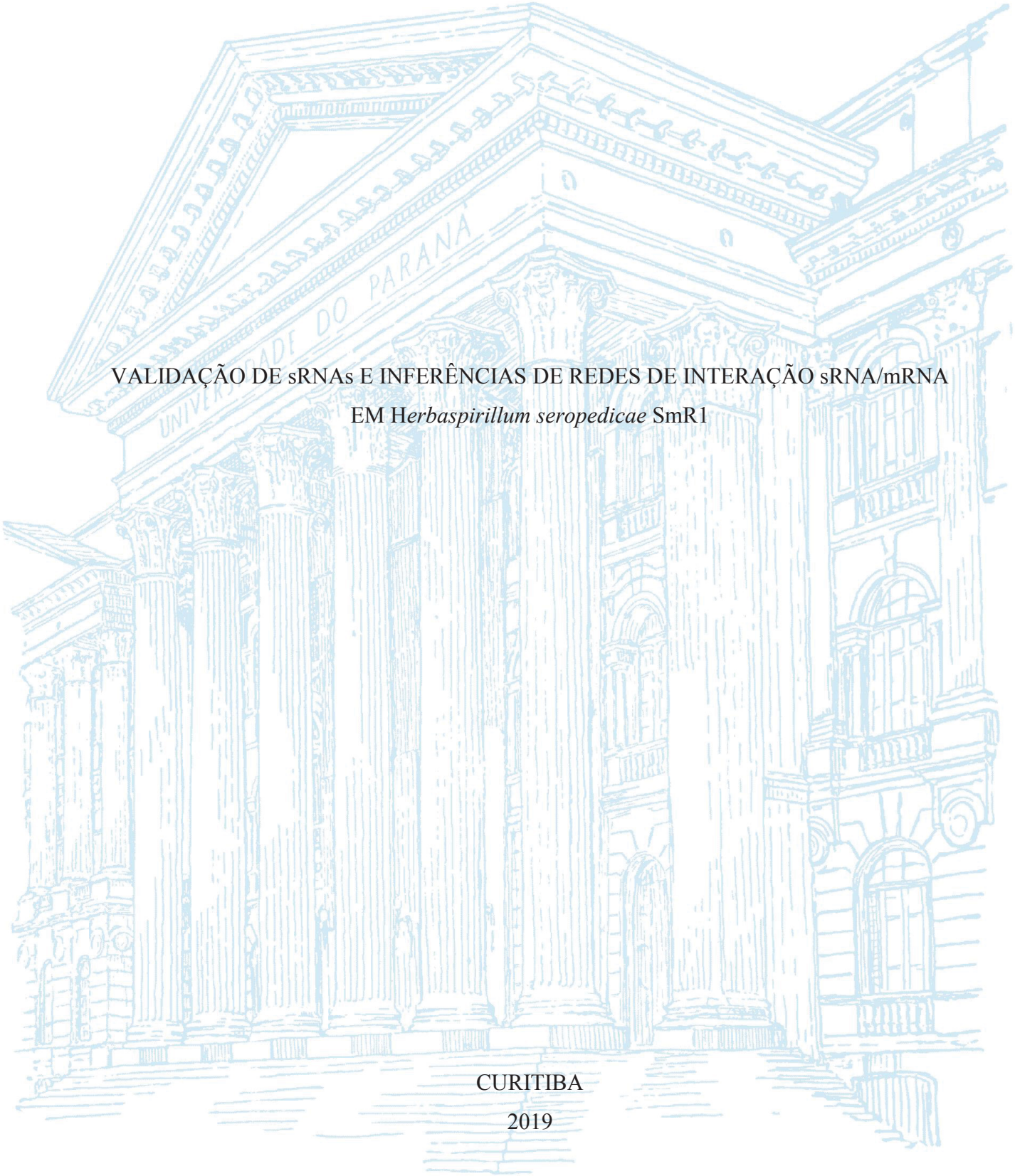
UNIVERSIDADE FEDERAL DO PARANÁ

TATIANE DOBRZANSKI



VALIDAÇÃO DE sRNAs E INFERÊNCIAS DE REDES DE INTERAÇÃO sRNA/mRNA EM H*erbaspirillum seropedicae* SmR1

CURITIBA

2019

TATIANE DOBRZANSKI

VALIDAÇÃO DE sRNAs E INFERÊNCIAS DE REDES DE INTERAÇÃO sRNA/mRNA EM H*erbaspirillum seropedicae* SmR1

Tese apresentada ao curso de Pós Graduação em Ciências – Bioquímica, Setor de Ciências Biológicas, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Doutor em Ciências – Bioquímica.

Orientadora: Prof[a]. Dr[a]. Maria Berenice Reynaud Steffens

Orientadora no Exterior: Dr[a]. Vânia Pobre

Co-orientador: Prof. Dr. Fabrício Martins Lopes

CURITIBA

2019

Universidade Federal do Paraná. Sistema de Bibliotecas.
Biblioteca de Ciências Biológicas.
(Giana Mara Seniski Silva – CRB/9 1406)

Dobrzanski, Tatiane
Validação de sRNAS e inferências de redes de interação sRNA/mRNA em Herbaspirillum seropedicae SmR1. / Tatiane Dobrzanski. – Curitiba, 2019.
109 p.: il.

Orientador: Maria Berenice Reynaud Steffens
Orientador no exterior: Vânia Pobre
Coorientador: Fabrício Martins Lopes

Tese (doutorado) - Universidade Federal do Paraná, Setor de Ciências Biológicas. Programa de Pós-Graduação em Ciências - Bioquímica.

1. Herbaspirillum 2. Nitrogênio - Fixação 3. RNA I. Título II. Steffens, Maria Berenice Reynaud III. Pobre, Vânia IV. Lopes, Fabrício Martins V. Universidade Federal do Paraná. Setor de Ciências Biológicas. Programa de Pós-Graduação em Ciências- Bioquímica.

CDD (22. ed.) 572.545

MINISTÉRIO DA EDUCAÇÃO
SETOR SETOR DE CIENCIAS BIOLOGICAS
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS
(BIOQUÍMICA) - 40001016003P2

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em CIÊNCIAS (BIOQUÍMICA) da Universidade Federal do Paraná foram convocados para realizar a arguição da tese de Doutorado de **TATIANE DOBRZANSKI** intitulada: **Validação de sRNAs e inferência de redes de interação sRNA/mRNA em *Herbaspirillum seropedicae* SmR1**, após terem inquirido a aluna e realizado a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de doutor está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 29 de Abril de 2019.

MARIA BERENICE REYNAUD STEFFENS
Presidente da Banca Examinadora (UFPR)

MAURO ANTONIO ALVES CASTRO
Avaliador Externo (UFPR)

EMANUEL MALTEMPI DE SOUZA
Avaliador Interno (UFPR)

ALEXANDRE ROSSI PASCHOAL
Avaliador Externo (UTFPR)

AGRADECIMENTOS

Agradeço à Prof[a]. Dra. Maria Berenice Reynaud Steffens pela sua dedicação e ensinamentos valiosos ao longo do doutorado que foram fundamentais para meu crescimento profissional e conclusão deste trabalho e pela confiança em conceder a mim a oportunidade de doutorado sanduíche. Ao Prof. Dr. Fabrício Lopes por todo ensinamento, dedicação e orientação. Ao Prof. Dr. Alexandre Paschoal por todo ensinamento e dedicação. À Prof[a]. Dra. Cecília Arraiano e à Dra. Vânia Pobre por me receberem com muito carinho no ITQB, pela paciência e excelente orientação, e por me concederem a oportunidade de permanecer por mais 6 meses em seu laboratório. Agradeço especialmente à Dra. Vânia Pobre por me ajudar em tudo o que eu precisava enquanto estava no exterior, por ser minha orientadora e amiga que jamais esquecerei. A todos os amigos de laboratório do ITQB e da Bioquímica pelos agradáveis momentos que ficarão marcados em minha memória para sempre.

Agradeço ao Prof. Dr. Emanuel Maltempi de Souza por toda dedicação e sugestões dadas para conclusão deste trabalho. Ao Núcleo de Fixação de Nitrogênio que forneceu os dados para desenvolvimento deste trabalho. Aos professores do Departamento de Bioquímica e Biologia Molecular e do PPG em Bioinformática que contribuíram imensamente para minha formação. À coordenação do PPG em Ciências-Bioquímica. Às agências CNPq, CAPES e FCT (Portugal) pelo auxílio financeiro.

Agradeço à minha família por toda paciência, amor e cuidado comigo enquanto estava longe. Ao meu namorado por todo amor, incentivo e por estar sempre comigo em todos os momentos. À Deus por iluminar meu caminho e guiar os meus passos em todos os momentos.

RESUMO

Pequenos RNAs não-codificadores (sRNAs) são reguladores chave no controle pós-transcricional da expressão gênica. Eles podem ser encontrados nos três domínios da vida e são particularmente importantes em bactérias, permitindo responder rapidamente as mudanças ambientais, modulando a rotatividade de mRNAs alvos e controlando a tradução. Os sRNAs podem ser divididos em três categorias, sRNAs codificados em *trans* (traRNA), sRNAs codificados em *cis* (caRNA) e *riboswitches*. *Herbaspirillum seropedicae* SmR1 é uma bactéria diazotrófica, essencialmente endofítica que pertence a classe das β-proteobactérias. Este microrganismo fixa nitrogênio sob condições microaeróbicas dentro dos tecidos vegetais de culturas de cereais economicamente importantes como trigo, arroz, milho e sorgo. O único cromossomo circular de *H. seropedicae* SmR1 já foi sequenciado. Desde então, muitos estudos focam na estrutura genômica, expressão gênica e fisiologia desta bactéria. Entretanto, uma investigação sobre a presença e função de sRNAs nunca foi realizada. O objetivo geral deste trabalho é validar a expressão de sRNAs e inferir redes de interação sRNAs-mRNAs de *H. seropedicae* SmR1 por meio de abordagens de bioinformáticas. Este trabalho foi dividido em dois capítulos. No primeiro capítulo, será apresentado a expressão de sRNAs no genoma de *H. seropedicae* SmR1 na presença de duas fontes diferentes de nitrogênio e / ou na presença de naringenina. Cento e dezessete sRNAs são expressos nesta bactéria nessas condições, porém, apenas 20 sRNAs apresentaram identidade de sequência com sRNAs bem caracterizados em outras bactérias conforme o banco de dados da Rfam. Nós conseguimos validar por *northern blot* a expressão de cinco sRNAs, Hsnc050 (RNA 6S), Hsnc028, Hsnc042, Hsnc073 e Hsnc082. No segundo capítulo, relata-se a inferência de redes de interação de sRNA-mRNA de *H. seropedicae* SmR1 baseados em duas abordagens para inferência de redes. Na primeira abordagem utilizamos dados de expressão (rede HsNetEx) e na segunda buscamos regiões de complementariedade de bases com mRNAs (rede HsNetEx). Quarenta e seis interações foram idênticas entre HsNetEx e HsNetBP. Uma rede de consenso que estabelece 3.218 relações envolvendo 77 traRNAs e 1.988 mRNAs foi obtida e confirma a multiplicidade de interações de traRNAs e seus alvos de mRNAs. Foram identificados diversos mRNAs alvos dos sRNAs analisados, dentre estes alguns envolvidos no metabolismo de nitrogênio. Esta abordagem de bioinformática é importante como primeiro passo para validação experimental de alvos e funções de traRNAs.

Keywords: *Herbaspirillum seropedicae* SmR1. sRNAs. RNA regulatório. fixação de nitrogênio. rede de sRNA-mRNA.

ABSTRACT

Small non-coding RNAs (sRNAs) have key regulatory roles in post-transcriptional control of gene expression. They can be found in all three domains of life and are particularly important in bacteria allowing them to rapidly respond to the environmental challenges and modulate turnover of target mRNAs and affect their translation. sRNAs can be divided in three categories *trans*-encoded RNAs (traRNA), *cis*-encoded RNAs (caRNA) and riboswitches. *Herbaspirillum seropedicae* SmR1 is a diazotrophic and obligate endophytic bacterium that belongs to the β-proteobacteria class. This microorganism fixes nitrogen under microaerobic conditions inside the tissues of the economically important cereal crops such as wheat, rice, maize and sorghum. The single circular chromosome of the *H. seropedicae* SmR1 strain was sequenced, since then, there are many studies focusing the genomic structure, gene expression and physiology of *H. seropedicae* SmR1. However, the investigation about the presence and function of sRNAs was never performed. The general objective of this work is to validate the expression of sRNAs and infer sRNAs-mRNA interaction networks *H. seropedicae* SmR1 based on bioinformatics approaches. This work was divided into two chapters. In the first chapter we reported the expression of several sRNA in *H. seropedicae* SmR1 genome in the presence of two nitrogen sources and/or in the presence of naringenin. One hundred and seventeen sRNAs were expressed in this bacterium under these conditions, but only 20 sRNAs had sequence identity with well-characterized sRNAs in another bacterium according to the Rfam database. We were able to validate by northern blot the expression of the Hsnc050 (6S RNA), Hsnc028, Hsnc042, Hsnc073 and Hsnc082 sRNAs. In the second chapter we reported the inference sRNA-mRNA interaction network in *H. seropedicae* SmR1 using two network inference approaches, using expression data (HsNetEx network) and searching for complementarity of bases with mRNAs (HsNetBP network). Forty-six interactions were identical between HsNetEx and HsNetBP. A consensus network which establishes 3,218 relationships involving 77 traRNAs and 1,988 mRNAs was obtained and confirms the complex interaction multiplicity of the traRNAs and their mRNAs targets. Several mRNAs were identified targeting the analysed sRNAs, among some involved in nitrogen metabolism. This bioinformatics approach represents an important step before the experimental validation of targets and functions of traRNAs.

Keywords: *Herbaspirillum seropedicae* SmR1. sRNAs. regulatory RNA. nitrogen fixation. sRNA-mRNA network.

SUMÁRIO

1. INTRODUÇÃO **10**

1.1. Pequenos RNAs 10

1.2. Predição de sRNAs 12

1.3. Redes de interação sRNA-mRNA 13

1.4. Redes baseadas em dados de expressão gênica 14

1.6. *Herbaspirillum seropedicae* SmR1 16

2. OBJETIVOS **18**

2.1. Objetivo geral 18

2.2. Objetivos específicos 18

4. ARTIGO 1: *In Silico* Prediction and Expression Profile Analysis of Small Non-coding RNAs in *Herbaspirillum seropedicae* SmR1 ..... **27**

4.1. ABSTRACT 27

4.2. INTRODUCTION 28

4.3. MATERIALS AND METHODS 30

4.3.1. Bacterial Growth **30**

4.3.2. Screening of small RNAs by nocoRNAc and Cufflinks **30**

4.3.3. Mapping and visualization of sequence reads and analyses of predicted sRNAs **31**

4.3.4. RNA -Seq data analysis **31**

4.3.5. RNA extraction and northern blot analysis **31**

4.4. RESULTS 33

4.4.1. sRNAs in the *H. seropedicae* SmR1 genome **33**

4.4.2. Identity of sRNAs of *H. seropedicae* SmR1 ..... **35**

4.4.3. Expression of sRNA in *H. seropedicae* SmR1 ..... **38**

4.4.4. Experimental validation of *H. seropedicae* SmR1 sRNAs **40**

4.5. DISCUSSION 42

4.6. CONCLUSIONS 46

4.7. REFERENCES 47

4.8. ADDITIONAL DATA 1. *Herbaspirillum seropedicae* SmR1 PREDICTED sRNAs.. 57

5. ARTIGO 2: NETWORK INTERACTIONS OF sRNAs-mRNAs FROM *Herbaspirillum seropedicae* SmR1 ..... **64**

5.2. INTRODUCTION 65

5.3. MATERIALS AND METHODS 67

5.3.1. RNA -Seq data analysis **67**

5.3.2. sRNA-mRNA network inferences based on expression data **67**

5.3.3. sRNA-mRNA network inferences based on base paring **67**

5.3.4. Consensus network inference **68**

5.4. RESULTS 70

5.4.1. sRNA-mRNA Network based on expression data **70**

5.4.2. The sRNAs targets of *H. seropedicae* SmR1 ..... **71**

5.4.3. sRNA-mRNA consensus network for *Herbaspirillum seropedicae* SmR1 ..... **72**

5.5. DISCUSSION 81

5.6. CONCLUSIONS 83

5.7. REFERENCES 85

5.8. ADDITIONAL DATA 92

5.8.1. Additional data 1: RNA -Seq data analysis **92**

5.8.2. Additional data 3: sRNA-mRNA common interactions between HsNetEx and the predictions of TargetRNA2 and IntaRNA. **93**

5.8.3. Additional data 4: Base pairing between sRNA-mRNA. **96**

6. CONCLUSÕES GERAIS **97**

7. REFERENCES **98**

1. INTRODUÇÃO

1.1. Pequenos RNAs

Os microrganismos necessitam se adaptar rapidamente às mudanças do ambiente para garantir sua sobrevivência em condições adversas. Para isso, a regulação da expressão gênica deve ser constante e rigorosa a fim de remodelar rapidamente o metabolismo celular (Carrier *et al.*, 2018). Pequenos RNAs (small RNAs - sRNAs) ou RNAs curtos não codificadores (small ncRNAs) são encontrados nos três domínios da vida e são particularmente importantes em bactérias permitindo a rápida adaptação ao ambiente (Dennis e Omer, 2005; Zaratiegui *et al.*, 2007). Os sRNAs são definidos como moléculas funcionais de RNAs com 50 a 500 nucleotídeos, capazes de modular o *turnover* e afetar a tradução de mRNAs alvos (Waters e Storz, 2009; Storz *et al.*, 2011). Eles são os principais reguladores pós-transcricionais em bactérias e arquéias e estão envolvidos em diversos processos celulares incluindo *quorum sensing*, virulência, metabolismo do carbono e respostas a condições de estresses (Melamed *et al.*, 2016; Nitzan *et al.*, 2017; Brosse e Guillier, 2018).

Os sRNAs possuem características regulatórias exclusivas como: (i) apresentar uma atividade catalítica ou serem componentes de ribonucleoproteínas (ex: porção do RNA da RNase P), (ii) podem interagir com proteínas afetando sua atividade (ex: 6S RNA, tmRNA), (iii) estar envolvidos na secreção de proteínas (ex: 4.5S) e (iv) regular mRNAs alvos de maneira pós-transcricional por meio de pareamento de bases (Brosse e Guillier, 2018). A maior parte dos sRNAs de bactérias encontra-se agrupada nesta última categoria.

Quanto à localização os sRNAs estão predominantemente localizados em regiões intergênicas (Storz *et al.*, 2011) e podem ser divididos em duas categorias principais: trans-encoded RNAs (traRNA) e cis-encoded RNAs (caRNA) (Zorgani *et al.*, 2016) de acordo com sua posição em relação ao seu alvo. Ambos os RNAs codificados *in cis* e *in trans* atuam por emparelhamento de bases com o mRNA alvo (Henkin, 2008).

Os caRNAs são transcritos a partir da fita de DNA complementar a seu alvo, geralmente se sobrepondo a extremidade 5' UTR do mRNA, e por isso apresentam uma extensa complementariedade de pares de bases com seu alvo, cerca de 75 nucleotídeos ou mais (Brantl, 2007; Waters e Storz, 2009). A formação do híbrido caRNA-mRNA pode bloquear a ligação ao ribossomo e inibir o processo de tradução, degradar ou clivar do mRNA

e, promover eventos de terminação traducional por meio da desestabilização do mRNA ao formar grampos no mRNA alvo (Caldelari *et al.*2013)

Os traRNAs são transcritos em regiões distantes de seu alvo e, por isso, presentam uma complementariedade de bases limitada em relação ao mesmo (Waters e Storz, 2009). Frequentemente requerem o auxílio da proteína Hfq, uma chaperona de RNA (Lenz *et al.* 2004, Melamed *et al.*, 2016) que ao formar o complexo sRNA:Hfq:mRNA pode atuar positiva ou negativamente na regulação pós-transcricional (Faner e Feig, 2013). O mecanismo clássico de ação dos traRNAs é o emparelhamento em regiões 5' UTR próximas ao sítio de ligação ao ribossomo (RBS) ou ao códon de início de tradução, porém pode haver também interação na sequência codificadora do mRNA alvo ou na região 3' UTR (Storz *et al.*, 2011). O pareamento sRNA/mRNA alvo pode ocasionar a degradação da cadeia dupla sRNA-mRNA, a clivagem do mRNA, a inibição ou ativação da tradução ou até mesmo a terminação da transcrição (Oliva *et al.*, 2015). As ribonucleases (RNAses), envolvidas no *turnover* de RNAs através da sua clivagem, frequentemente estão associadas na clivagem e degradação dos sRNAs (Saramago *et al.* 2014). A inibição da tradução pode ocorrer mais precisamente se houver o pareamento do sRNA a montante do códon de início de tradução, neste caso, o ribossomo não é capaz de se ligar a sequência RBS. Já foi observado que alguns sRNAs emparelham na ORF de mRNAs alvos, o que não impede o início da tradução, porém desestabiliza o transcrito mRNA e pode ocorrer o recrutamento do degradossoma de RNA, acarretando na clivagem do mRNA pela RNase E (Carrier *et al.*, 2018). Existem ainda aqueles traRNAS que ao se ligarem a proteínas modulam a sua atividade, como é o caso do 6S RNA (Wassarman e Storz, 2000).

Bactérias também apresentam *riboswitches* que são sequências de bases localizadas na região 5'UTR de mRNAs e constituem sítios de ligação de metabólitos intracelulares ou íons metálicos para regulação da transcrição desses mRNAs (Waters e Storz, 2009, Serganov e Nudler, 2013, Hallberg *et al.*, 2017). *Riboswitches* são formados por um aptâmero e uma plataforma de expressão (Breaker, 2018). O aptâmento é um domínio altamente conversado em sequência e estrutura entre diversos organismos e a região de ligação do ligante. A plataforma de expressão fica imediatamente a jusante do aptâmero, sendo variável em seqüência, estrutura e tamanho, que permite a regulação do mRNA por meio da alteração na estrutura na porção 5'UTR do mRNA (Serganov e Nudler, 2013; Mellin e Cossart, 2015; BREAKER, 2018). O modo de ação dos *riboswitches* inicia-se com a ligação do metabólito alvo ao aptâmero. Esta ligação resulta em mudanças conformacionais na plataforma de

expressão que regula o processo de tradução do mRNA (Serganov e Nudler, 2013; Mellin e Cossart, 2015; Breaker, 2018).

Um tipo particular de *riboswitch*, o sensorial ou também como conhecido como termômetro de RNA, também pode ser encontrado em bactéria (Waters e Storz, 2009). Ao contrário dos *riboswitches* de ligação à metabólitos, alteram a conformação dos RNAs regulados em resposta à alteração da temperatura intracelular, sendo por isso, chamados de termômetros de RNA (Narberhaus *et al.*, 2005).

1.2. Predição de sRNAs

O sRNA 6S RNA ou ssrS de *Escherichia coli* foi o primeiro exemplo estudado de sRNA, por meio de géis de poliacrilamida (Hindley, 1967) e o primeiro RNA sequenciado (Brownlee, 1971). Entretanto, esse sRNA só foi considerado como um RNA regulatório após a descoberta do sRNA MicF em *E. coli* por Mizuno e colaboradores (1984) que então estabeleceram o conceito de genes regulados por RNAs (Wong *et al.*, 2018). O MicF foi o primeiro sRNA caracterizado por inibir a tradução do mRNA *ompF*, que codifica uma porina da membrana externa e impede a entrada de moléculas deletérias na célula (Delihas e FORST, 2001; Delihas, 2015).

O aprimoramento e a aplicação de metodologias laboratoriais de busca de sRNA como eletroforese uni e bidimensional, *northern blotting*, *microarray* e sequenciamento de DNA e de RNA permitiram ampliar o número de sRNAs identificados em diferentes microrganismos (Vogel e Sharma, 2005; Liu *et al.*, 2010; NawaZ *et al.*, 2017). Em paralelo, com o desenvolvimento da bioinformática e à disponibilidade de genomas em bancos de dados, diferentes softwares para predição *in silico* de sRNAs foram desenvolvidos com a finalidade de descobrir sequências candidatas a serem sRNAs em um maior número de microrganismos.

A maioria dos sRNAs conhecidos estão disponíveis no banco de dados de famílias de RNA, o Rfam (http://rfam.org/), onde cada família é representada por uma alinhamento múltiplo de sequência, uma estrutura secundária consenso e um modelo de covariância (Kalvari *et al.*, 2017; Kalvari *et al.*, 2018). Por meio do Rfam é possível comparar possíveis sequências de sRNAs com sRNAs já caracterizados e assim, determinar uma identidade para sRNAs (). Em janeiro/2019 esse banco possuía 3.016 famílias de RNAs.

Os primeiros softwares para predições de sRNAs baseavam-se principalmente em pesquisas de conservação de sequências de regiões intergênicas (IGRs) entre bactérias (Rivas

et al., 2001; Wassarman *et al.*, 2001; Altuvia *et al.*, 2007). Posteriormente, a ferramentas foram desenvolvidas para buscar um conjunto de características para predizer um sRNA, como a estrutura secundária da molécula, estabilidade termodinâmica, conservação de sequência, sinais de terminação transcricional independentes de Rho e agrupamentos de sequências não codificantes (Vogel e Sharma, 2005; Lu *et al.*, 2011; Sridhar e Gunasekaran, 2013; Nawaz *et al.*, 2017). Dentre as ferramentas de predição de RNAs não codificadores podemos citar o nocoRNAc (no-coding RNA characterization), um programa que considera a busca por sinais terminadores independentes Rho e amplia a busca por sRNAs em regiões codificadoras com o objetivo de prever RNA codificados em cis (Herbig e Nieselt, 2011). Outro exemplo é o Infernal (Inference of RNA Alignment) que utiliza modelo de covariância para busca por homologias de RNA em banco de dados de famílias de RNA, incluindo homologia de estrutura secundária (Nawrocki e Eddy, 2013).

Após a predição de sRNAs por softwares bioinformática é essencial investigar se eles estão sendo expressos nas células. Na última década, o sequenciamento de DNA complementar em larga escala (RNA-seq) tem facilitado a observação da expressão gênica desse tipo de moléculas em diversos organismos (Lopes, 2011; Marbach *et al.*, 2012). Pelo RNA-seq pode-se quantificar a expressão de mRNAs e sRNAs em condições experimentais específicas e detectar até mesmo genes fracamente expressos com um alto nível de precisão e profundidade (Marguerat e Bähler, 2010; Schulze *et al.*, 2015). Essa metodologia possibilitou a descoberta em estudo de diversos sRNAs como nos trabalhos com *Streptococcus pyogenes* (Le Rhun *et al.*, 2016), *Pseudomonas aeruginosa* (Gómez-Lozano *et al.*, 2012), *Listeria monocytogenes* (Mellin e Cossart, 2012) e *Vibrio cholerae* (Papenfort *et al.*, 2015). Uma ferramenta usada para localizar transcritos em regiões intergênicas, inclusive transcritos de sRNAs é a Cufflinks (Trapnell *et al.*, 2010).

1.3. Redes de interação sRNA-mRNA

A interação sRNA-mRNA pode modular a tradução por meio da modificação da estrutura secundária do mRNA, da estabilidade do transcrito ou da interação com ribossomos e com proteínas que se ligam ao mRNA. O avanço dos estudos envolvendo sRNAs regulatórios demonstraram que um dado sRNA pode regular a tradução de um ou de vários mRNAs podendo formar uma rede de regulação (Zhan *et al.*, 2016; Beisel e Storz, 2010).

A medição simultânea dos níveis de expressão de milhares genes, sob diferentes condições experimentais, gera conjuntos de dados que podem fornecer informações sobre as funções específicas de genes já anotados, genes hipotéticos e dos sRNAs. A integração desses dados com o objetivo de obter informações relevantes a respeito da fisiologia global do organismo estudado é um desafio que a biologia de sistemasvem superando por meio da inferência de redes de interação gênica (Hecker *et al.*, 2009). Por meio dessa abordagem um organismo é considerado uma rede complexa de moléculas interconectadas por uma série de processos (replicação, transcrição, tradução e outras vias metabólicas) e, dessa forma, pode-se entender o funcionamento desse organismo de forma holística (Lopes *et al.*, 2014).

A inferência de redes de interação utiliza-se de técnicas de engenharia reversa, na qual, a partir de dados de expressão gênica (dados finais) inferem-se interações entre os genes de uma rede, indicando como ocorre a regulação gênica (o início) (Hecker *et al.*, 2009).

Em um genoma bacteriano típico, cerca de 100 a 200 sRNAs estão presentes e cada sRNA pode regular um ou dezenas de alvos, tornando a rede de interação sRNAs-mRNA uma parte significativa da rede global de regulação gênica (Pain *et al.*, 2015). Apesar da grande importância dos sRNAs, poucos trabalhos ainda se propõem a compreender sua função em um sistema mais amplo como uma rede de interações em um organismo.

Alguns trabalhos relatam com sucesso em estabelecer redes de regulação gênica com dados públicos de transcriptoma e/ou softwares de predição de alvos de mRNAs, apresentando importantes interações dos genes, até então desconhecidas (Modi *et al.*, 2011; Skippington e Ragan, 2012; Zhang *et al.*, 2012; Zhu *et al.*, 2014). De maneira geral, redes de interação sRNAs-mRNAs podem ser construídas por métodos baseados em dados de expressão gênica e em pareamento de bases.

1.4. Redes baseadas em dados de expressão gênica

Diversos métodos tem sido desenvolvidos para inferência de redes de interação baseados em dados de expressão, dentre eles podemos citar o *Algorithm for the Reconstruction of Accurate Cellular Networks* (Aracne) (Margolin *et al.*, 2006), *Context Likelihood Relatedness* (CLR) (Faith *et al.*, 2007), *Minimum Redundancy / Maximum Relevance Networks* (MRNET) (Meyer *et al.*, 2008), *Gene Network Inference with Ensemble of Trees* (GENIE3) (Irrthum *et al.*, 2010), e *DimReduction - Sequential Forward Floating Selec* (SFFS) (Lopes *et al.*, 2008).

As redes inferidas com ARACNE, CLR e MRNET são baseadas em informação mútua. Inicialmente, a estimativa da informação mútua é realizada entre todas as variáveis. Posteriormente, um filtro é executado para selecionar os relacionamentos relevantes. A informação mútua (MI) é uma medida de informação compartilhada entre duas variáveis que são amostradas aleatoriamente (Battiti, 1994; Lopes *et al.*, 2009). Posteriormente, um filtro é executado para selecionar os relacionamentos relevantes.O GENIE3 infere uma rede de genes baseada na seleção de variáveis com *ensembles* de árvores de regressão. Esse método inicialmente divide a predição da rede em diferentes problemas de regressão. Usando métodos baseados em árvores, florestas aleatórias ou árvores extras, cada problema de regressão é resolvido encontrando um padrão de expressão de um gene alvo a partir do padrão de expressão de outros genes de entrada (Irrthum *et al.*, 2010). A principal característica da rede de DimReduction é a possibilidade de inferência de rede com propriedades globais, mesmo se houver uma quantidade limitada de amostras de dados de expressão. Ao contrário das redes mencionadas anteriormente que inferem interações par-a-par, DimReduction procura um subconjunto de genes alvo que expliquem o comportamento de um gene (Lopes *et al.*, 2008).

1.5. Redes por pareamento de bases

A predição de regiões de pareamento de bases é uma importante abordagem para pesquisa de alvos entre sRNAs-mRNAs. Os traRNAs interagem com seus alvos através de pequenas regiões para formação de um sRNA-mRNA duplex sendo que apenas 6-10 nucleotídeos são necessários. Esta curta sequência necessária para formação do duplex é um dos requisitos analisados para predição computacional de alvos de sRNA (Wong et., 2018).

As ferramentas targetRNA2 e IntaRNA pesquisam regiões de pareamento de bases de sRNAs próximas ao sítio de ligação ao ribossomo (RBS) (Kery *et al.*, 2014; Busch *et al.*, 2008). Um conjunto de recursos é usado para predizer alvos de sRNA com essas ferramentas. A ferramenta TargetRNA2 considera a conservação da seqüência de sRNA em outras bactérias, a acessibilidade de regiões na estrutura do sRNA secundário, bem como os mRNAs alvo e a energia de hibridação entre o mRNA alvo e o sRNA (Kery *et al.*, 2014). A ferramenta IntaRNA prevê interações sRNA-mRNA próximas a região do RBS de cada gene considerando a acessibilidade dos locais do mRNA para a interação com o sRNA e uma sequência com no mínimo 7 nucleotídeos necessária para iniciar a interação com o sRNA IntaRNA calcula um escore de energia, que representa a energia livre de hibridação das

subsequências que interagem, e as energias livres necessárias para desdobrar os locais de interação em ambas as moléculas de RNA para torná-la acessível durante a interação sRNA-mRNA (Busch *et al.*, 2008; Mann *et al.*, 2017).

1.6. *Herbaspirillum seropedicae* SmR1

Herbaspirillum seropedicae SmR1 é uma bactéria da Classe β-proteobacteria, endofítica, não patogênica, e diazotrófica que coloniza espaços intercelulares de monocotiledôneas e dicotiledôneas, incluindo culturas economicamente importantes como cana-de-açúcar, milho, trigo, arroz e banana (Baldani *et al.*, 1986; Monteiro *et al*, 1999; Cruz *et al.*, 2001; Roncato-Maccari *et al.*, 2003; Baldani e Baldani, 2005). A estirpe SmR1 é um mutante espontâneo resistente a estreptomicina (100 μg/mL) isolado de *H. seropedicae* Z78 (Baldani *et al.*, 1986, Souza *et al.*, 2000). *In planta,* o ambiente microaeróbico favorece a expressão e atividade dos genes do operon *nifHDK* e outros genes *nif* em *H. seropedicae* SmR1, que fixa nitrogênio (Roncato-Maccari *et al.*, 2003). A associação *Herbaspirillum*-planta promove um melhor desenvolvimento do vegetal que se beneficia de parte do nitrogênio fixado da biossíntese de fitohormônios.

O genoma de *H. seropedicae* SmR1 foi sequenciado pelo Consórcio Genopar (Pedrosa *et al.*, 2011) e foram encontrados 4.804 genes. Entretanto, nenhum RNA curto não codificador foi identificado. A disponibilidade da sequência genômica permitiu o avanço de estudos relacionados à estrutura genômica, expressão gênica e fisiologia dessa bactéria (Weiss, 2010; Marin *et al.*, 2013; Cardoso, 2015; Tadra-Sfeir *et al.*, 2015; Balsanelli *et al.*, 2016; Bonato *et al.*, 2016; Pankievicz *et al.*, 2016; Batista *et al.*, 2018).

Em paralelo, a investigação da existência e função dos RNA curtos não codificadores (sRNAs) iniciou com pela abordagem bioinformática visando a predição de candidatos. Projetos de dissertação de mestrado desenvolvidos por Moreno, L.F. (2013) e Barbosa, H.C.S. (2014) resultaram em um conjunto de 228 sequências de RNAs candidatas a serem sRNAs em *H. seropedicae* SmR1. Nessas predições foram utilizadas as ferramentas nocoRNAc (Herbig *et al.*, 2011), CuffLinks (Trapnell *et al.*, 2010) e Infernal 1.1.1 (Nawrocki e Eddy, 2013), sendo que somente essa última fornece informação sobre uma provável identidade do sRNA.

Uma vez que no Núcleo de Fixação Nitrogênio foram desenvolvidos diversos projetos de pesquisa que resultaram na produção de um banco dados de sequenciamento de RNA (RNA-seq) de *H. seropedicae* SmR1 e que já se dispunha de um conjunto de candidatos a

sRNA, foi iniciado um projeto de pesquisa com o objetivo de aprofundar a análise dos sRNA candidatos e avançar na pesquisa relacionada à determinação da identidade e de prováveis alvos dos trans sRNA. Os resultados estão apresentados em dois capítulos, sendo o primeiro relativo à validação dos sRNA anteriormente preditos e o segundo relativo à inferência de redes interação traRNAs-mRNAs.

2. OBJETIVOS

2.1. Objetivo geral

Validar a expressão de sRNAs e inferir redes de interação de sRNA-mRNA em *Herbaspirillum. seropedicae* SmR1.

2.2. Objetivos específicos

- Analisar a expressão dos sRNAs em dados de RNA-seq disponíveis no Núcleo de Fixação Biológica de Nitrogênio.

- Determinar a identidade dos sRNAs.

- Validar a expressão de sRNAs por meio *de northern blot*.

- Inferir redes de interação traRNAs-mRNAs utilizando métodos de inferência de rede baseados em dados de expressão de RNA-seq.

- Pesquisar mRNAs alvos utilizando ferramentas de bioinformática baseadas no pareamento de bases traRNA-mRNA.

- Desenvolver uma abordagem para inferência de redes de regulação de interação traRNA-mRNA em procariotos.

3. REFERÊNCIAS

Altuvia, S. (2007). Identification of bacterial small non-coding RNAs: experimental approaches. *Current opinion in microbiology* 10(3), 257-261.

Baldani, J.I., and Baldani, V.L.D. (2005). History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. *Anais da Academia Brasileira de Ciências* 77, 549-579.

Baldani, J.I., Pot, B., Kirchhof, G., Falsen, E., Baldani, V.L., Olivares, F.L., *et al.* (1996). Emended description of *Herbaspirillum*; inclusion of [*Pseudomonas*] rubrisubalbicans, a milk plant pathogen, as *Herbaspirillum rubrisubalbicans* comb. nov.; and classification of a group of clinical isolates (EF group 1) as *Herbaspirillum* species 3. *Int J Syst Bacteriol* 46(3), 802-810. doi: 10.1099/00207713-46-3-802.

Barbosa, H. C. D. S. (2014). Re-anotação do genoma de *Herbaspirillum seropedicae* SmR1 com dados de transcriptoma por RNA_Seq.

Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nature genetics* 37(4), 382. doi: 10.1038/ng1532.

Batista, M.B., Teixeira, C.S., Sfeir, M.Z.T., Alves, L.P.S., Valdameri, G., Pedrosa, F.O., *et al.* (2018). PHB Biosynthesis Counteracts Redox Stress in *Herbaspirillum seropedicae*. *Front Microbiol* 9, 472. doi: 10.3389/fmicb.2018.00472.

Battiti, R. (1994). Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on neural networks* 5(4), 537-550.

Bonato, P., Batista, M.B., Camilios-Neto, D., Pankievicz, V.C.S., Tadra-Sfeir, M.Z., Monteiro, R.A., *et al.* (2016). RNA-seq analyses reveal insights into the function of respiratory nitrate reductase of the diazotroph *Herbaspirillum seropedicae*. *Environmental microbiology* 18(8), 2677--2688. doi: 10.1111/1462-2920.13422.

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Current opinion in microbiology* 10(2), 102-109.

Beisel, C. L., and Storz, G. (2010). Base pairing small RNAs and their roles in global regulatory networks. *FEMS microbiology reviews* 34(5), 866-882.

Breaker, R. R. (2018). Riboswitches and translation control. *Cold Spring Harbor perspectives in biology* 10(11), a032797.

Brosse, A., and Guillier, M. (2018). Bacterial Small RNAs in Mixed Regulatory Networks. *Microbiology spectrum* 6(3).

Brownlee, G. G. (1971). Sequence of 6S RNA of *E. coli*. *Nature New Biology* 229(5), 147.

Busch, A., Richter, A. S., and Backofen, R. (2008). IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24(24), 2849-2856. doi:10.1093/bioinformatics/btn544.

Caldelari, I., Chao, Y., Romby, P., and Vogel, J. (2013). RNA-mediated regulation in pathogenic bacteria. *Cold Spring Harbor perspectives in medicine* 3(9), a010298.

Cardoso, R. L. A. (2015). Análise genômica comparativa de bactérias do gênero *Herbaspirillum*.

Carrier, M. C., Lalaouna, D., and Massé, E. (2018). Broadening the definition of bacterial small RNAs: Characteristics and mechanisms of action. *Annual review of microbiology* 72, 141-161.

Coleman, J., Green, P. J., and Inouye, M. (1984). The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. *Cell* 37(2), 429-436.

Cruz, L. M., de Souza, E. M., Weber, O. B., Baldani, J. I., Döbereiner, J., and de Oliveira Pedrosa, F. (2001). 16S ribosomal DNA characterization of nitrogen-fixing bacteria isolated from banana (Musa spp.) and pineapple (Ananas comosus (L.) Merril). *Appl. Environ. Microbiol*., 67(5), 2375-2379. doi: 10.1128/AEM.67.5.2375-2379.2001.

Delihas, N. (2015). Discovery and characterization of the first non-coding RNA that regulates gene expression, micF RNA: A historical perspective. *World journal of biological chemistry* 6(4), 272.

Delihas, N., and Forst, S. (2001). MicF: an antisense RNA gene involved in response of *Escherichia coli* to global stress factors. *Journal of molecular biology* 313(1), 1-12.

Dennis, P. P., and Omer, A. (2005). Small non-coding RNAs in Archaea. *Current opinion in microbiology* 8(6), 685-694.

Melamed, S., Peer, A., Faigenbaum-Romm, R., Gatt, Y. E., Reiss, N., Bar, A., *et al.* (2016). Global mapping of small RNA-target interactions in bacteria. *Molecular cell* 63(5), 884-897.

Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., *et al.* (2007). Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS biology* 5(1), e8. doi:10.1371/journal.pbio.0050008.

Faner, M. A., and Feig, A. L. (2013). Identifying and characterizing Hfq–RNA interactions. *Methods* 63(2), 144-159.

Gómez-Lozano, M., Marvig, R. L., Molin, S., and Long, K. S. (2012). Genome-wide identification of novel small RNAs in *Pseudomonas aeruginosa*. *Environmental microbiology* 14(8), 2006-2016. doi: 10.1111/j.1462-2920.2012.02759.x.

Hallberg, Z. F., Su, Y., Kitto, R. Z., and Hammond, M. C. (2017). Engineering and in vivo applications of riboswitches. *Annual review of biochemistry* 86, 515-539.

Hecker, M., Lambeck, S., Toepfer, S., Van Someren, E., and Guthke, R. (2009). Gene regulatory network inference: data integration in dynamic models-a review. *Biosystems* 96(1), 86-103.

Henkin, T. M. (2008). Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes and development* 22(24), 3383-3390.

Herbig, A., and Nieselt, K. (2011). nocoRNAc: characterization of non-coding RNAs in prokaryotes. *BMC bioinformatics* 12(1), 40.

Hindley, J. T. (1967). Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. *Journal of molecular biology* 30(1), 125-136.

Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PloS one* 5(9), e12776. doi:10.1371/journal.pone.0012776.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E. P., Rivas, E., Eddy, S. R., *et al.* (2017). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic acids research* 46(D1), D335-D342.

Kalvari, I., Nawrocki, E. P., Argasinska, J., Quinones-Olvera, N., Finn, R. D., Bateman, A., and Petrov, A. I. (2018). Non-Coding RNA Analysis Using the Rfam Database. *Current protocols in bioinformatics* 62(1), e51.

Kery, M. B., Feldman, M., Livny, J., and Tjaden, B. (2014). TargetRNA2: identifying targets of small regulatory RNAs in bacteria. *Nucleic acids research* 42(W1), W124-W129. doi: 10.1093/nar/gku317.

Le Rhun, A., Beer, Y. Y., Reimegård, J., Chylinski, K., and Charpentier, E. (2016). RNA sequencing uncovers antisense RNAs and novel small RNAs in *Streptococcus pyogenes*. *RNA biology* 13(2), 177-195.

Lenz, D. H., Mok, K. C., Lilley, B. N., Kulkarni, R. V., Wingreen, N. S., and Bassler, B. L. (2004). The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. Cell 118(1), 69-82.

Liu, J. M.; Camilli, A. A broadening world of bacterial small RNAs. *Curr Opin Microbiol* v. 13, n. 1, p. 18-23, 2010.

Lopes, F. M., Cesar Jr, R. M., and Costa, L. D. F. (2011). Gene expression complex networks: synthesis, identification, and analysis. *Journal of Computational Biology* 18(10), 1353-1367. doi: 10.1089/cmb.2010.0118.

Lopes, F.M., Martins, D.C., and Cesar, R.M. (2008). Feature selection environment for genomic applications. *BMC bioinformatics* 9(1), 451. doi: 10.1186/1471-2105-9-451.

Lopes, F. M., Martins, D. C., and Cesar, R. M. (2009, May). Comparative study of GRNs inference methods based on feature selection by mutual information. In 2009 IEEE International *Workshop on Genomic Signal Processing and Statistics* (pp. 1-4). IEEE. doi: 10.1109/GENSIPS.2009.5174334.

Lopes, F. M., Martins Jr, D. C., Barrera, J., and Cesar Jr, R. M. (2014). A feature selection technique for inference of graphs from their known topological properties: Revealing scale-free gene regulatory networks. *Information Sciences* 272, 1-15. doi: 10.1016/j.ins.2014.02.096.

Lu, X., Goodrich-Blair, H., and Tjaden, B. (2011). Assessing computational tools for the discovery of small RNA genes in bacteria. *RNA* 17(9), 1635-1647. doi:10.1261/rna.2689811.

Mann, M., Wright, P. R., and Backofen, R. (2017). IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic acids research* 45(W1), W435-W439. doi: 10.1093/nar/gkx279

Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., *et al.* (2012). Wisdom of crowds for robust gene network inference. *Nature methods* 9(8), 796.

Marin, A.M., Souza, E.M., Pedrosa, F.O., Souza, L.M., Sassaki, G.L., Baura, V.A., *et al.* (2013). Naringenin degradation by the endophytic diazotroph *Herbaspirillum seropedicae* SmR1. *Microbiology* 159(Pt 1), 167-175. doi: 10.1099/mic.0.061135-0.

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006, March). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In BMC

bioinformatics (Vol. 7, No. 1, p. S7). *BioMed Central.* doi:10.1186/1471-2105-7-S1-S7.

Marguerat, S., and Bähler, J. (2010). RNA-seq: from technology to biology. *Cellular and molecular life sciences* 67(4), 569-579.

Mellin, J. R., and Cossart, P. (2012). The non-coding RNA world of the bacterial pathogen *Listeria monocytogenes. RNA biology* 9(4), 372-378.

Mellin, J. R., and Cossart, P. (2015). Unexpected versatility in bacterial riboswitches. *Trends in Genetics* 31(3), 150-156.

Meyer, P. E., Lafitte, F., and Bontempi, G. (2008). minet: AR/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC bioinformatics* 9(1), 461. doi:10.1186/1471-2105-9-461.

Modi, S. R., Camacho, D. M., Kohanski, M. A., Walker, G. C., and Collins, J. J. (2011). Functional characterization of bacterial sRNAs using a network biology approach. *Proceedings of the National Academy of Sciences* 108(37), 15522-15527.

Monteiro, R. A., Souza, E. M., Funayama, S., Yates, M. G., Pedrosa, F. O., & Chubatsu, L. S. (1999). Expression and functional analysis of an N-truncated NifA protein of *Herbaspirillum seropedicae*. *FEBS letters* 447(2-3), 283-286.

Moreno, L. F. (2013). Predição e análise da expressão de RNAS não codificadores com função regulatória presentes na bactéria *Herbaspirillum Seropedicae* SmR1.

Narberhaus, F., Waldminghaus, T., and Chowdhury, S. (2006). RNA thermometers. *FEMS microbiology reviews* 30(1), 3-16.

Nawaz, M. Z., Jian, H., He, Y., Xiong, L., Xiao, X., and Wang, F. (2017). Genome-wide detection of small regulatory RNAs in deep-sea bacterium *Shewanella piezotolerans* WP3. *Frontiers in Microbiology* 8, 1093. https://doi.org/10.3389/fmicb.2017.01093

Nawrocki, E. P., and Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29(22), 2933-2935.

Nitzan, M., Rehani, R., and Margalit, H. (2017). Integration of bacterial small RNAs in regulatory networks. *Annual review of biophysics* 46, 131-148.

Oliva, G., Sahr, T., and Buchrieser, C. (2015). Small RNAs, 5′ UTR elements and RNA-binding proteins in intracellular bacteria: impact on metabolism and virulence. *FEMS microbiology reviews* 39(3), 331-349. doi: 10.1093/femsre/fuv022.

Pain, A., Ott, A., Amine, H., Rochat, T., Bouloc, P., and Gautheret, D. (2015). An assessment of bacterial small RNA target prediction programs. *RNA biology* 12(5), 509-513. doi: 10.1080/15476286.2015.1020269.

Pankievicz V.C., Camilios-Neto D., Bonato P., Balsanelli E., Tadra-Sfeir M.Z., Faoro H., Chubatsu L.S., Donatti L., Wajnberg G., Passetti F., Monteiro R.A., Pedrosa F.O., and Souza E.M. (2016). RNA-seq transcriptional profiling of *Herbaspirillum seropedicae* colonizing wheat (Triticum aestivum) roots. *Plant Mol Biol* 90(6), 589-603. doi: 10.1007/s11103-016-0430-6.

Papenfort, K., Förstner, K. U., Cong, J. P., Sharma, C. M., and Bassler, B. L. (2015). Differential RNA-seq of *Vibrio cholerae* identifies the VqmR small RNA as a regulator of biofilm formation. *Proceedings of the National Academy of Sciences* 112(7), E766-E775.

Pedrosa, F. O., Monteiro, R. A., Wassem, R., Cruz, L. M., Ayub, R. A., Colauto, N. B., *et al.* (2011). Genome of *Herbaspirillum seropedicae* strain SmR1, a specialized diazotrophic endophyte of tropical grasses. *PLoS genetics* 7(5), e1002064. https://doi.org/10.1371/journal.pgen.1002064

Rivas, E., and Eddy, S. R. (2001). Noncoding RNA gene detection using comparative sequence analysis. *BMC bioinformatics* 2(1), 8.

Roncato-Maccari, L. D., Ramos, H. J., Pedrosa, F. O., Alquini, Y., Chubatsu, L. S., Yates, M. G., *et al.* (2003). Endophytic *Herbaspirillum seropedicae* expresses nif genes in gramineous plants. *FEMS microbiology ecology* 45(1), 39-47. doi.org/10.1016/S0168-6496(03)00108-9.

Saramago, M., Bárria, C., dos Santos, R.F., Silva, I.J., Pobre, V., Domingues, S., *et al.* (2014). The role of RNases in the regulation of small RNAs. *Current opinion in microbiology* 18, 105-115. doi: 10.1016/j.mib.2014.02.009.

Schultze, T., Hilker, R., Mannala, G. K., Gentil, K., Weigel, M., Farmani, N., *et al.* (2015). A detailed view of the intracellular transcriptome of *Listeria monocytogenes* in murine macrophages using RNA-seq. *Frontiers in microbiology* 6, 1199. doi: 10.3389/fmicb.2015.01199.

Serganov, A., and Nudler, E. (2013). A decade of riboswitches. Cell, 152(1-2), 17-24.

Skippington, E., and Ragan, M. A. (2012). Evolutionary dynamics of small RNAs in 27 *Escherichia coli* and *Shigella genomes*. Genome biology and evolution 4(3), 330-345.

Souza, E. M., Pedrosa, F. O., Rigo, L. U., Machado, H. B., and Yates, M. G. (2000). Expression of the nifA gene of *Herbaspirillum seropedicae*: role of the NtrC and NifA binding sites and of the− 24/− 12 promoter element. *Microbiology* 146(6), 1407-1418. doi: 10.1099/00221287-146-6-1407.

Sridhar, J., and Gunasekaran, P. (2013). Computational small RNA prediction in bacteria. *Bioinformatics and biology insights* 7, BBI-S11213. https://doi.org/10.4137/BBI.S11213

Storz, G., Vogel, J., and Wassarman, K. M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. *Molecular cell* 43(6), 880-891.

Tadra-Sfeir, M.Z., Faoro, H., Camilios-Neto, D., Brusamarello-Santos, L., Balsanelli, E., Weiss, V., *et al.* (2015). Genome wide transcriptional profiling of *Herbaspirillum seropedicae* SmR1 grown in the presence of naringenin. *Frontiers in Microbiology* 6(MAY), 1-8. doi: 10.3389/fmicb.2015.00491.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., *et al.* (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7(3), 562--578. doi: 10.1038/nprot.2012.016.

Vogel, J., and Sharma, C. M. (2005). How to find small non-coding RNAs in bacteria. *Biological chemistry* 386(12), 1219-1238.

Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. *Genes and development* 15(13), 1637-1651.

Wassarman, K. M., and Storz, G. (2000). 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* 101(6), 613-623.

Waters, L. S., and Storz, G. (2009). Regulatory RNAs in bacteria. Cell, 136(4), 615-628.

Weiss, V. A. (2010). Estratégias de finalização da montagem do genoma da Bactéria Diazotrófica Endofítica *Herbaspirillum Seropedicae* SmR1.

Wong, J., Pang, I., Wilkins, M., and Tree, J. J. (2018). Systems-Level Analysis of Bacterial Regulatory Small RNA Networks. *In Systems Biology* (pp. 97-127). Springer, Cham.

Zhan, S., Dong, Y., Zhao, W., Guo, J., Zhong, T., Wang, L., *et al.* (2016). Genome-wide identification and characterization of long non-coding RNAs in developmental skeletal muscle of fetal goat. *BMC genomics* 17(1), 666. doi: 10.1186/s12864-016-3009-3.

Zhang, J., Lu, K., Xiang, Y., Islam, M., Kotian, S., Kais, Z., *et al.* (2012). Weighted frequent gene co-expression network mining to identify genes involved in genome

stability. *PLoS computational biology* 8(8), e1002656. https://doi.org/10.1371/journal.pcbi.1002656

Zhu, P., and Han, J. (2014). Asynchronous stochastic Boolean networks as gene network models. *Journal of Computational Biology*, 21(10), 771-783. doi.org/10.1089/cmb.2014.0057

Zaratiegui, M., Irvine, D. V., and Martienssen, R. A. (2007). Noncoding RNAs and gene silencing. *Cell* 128(4), 763-776.

Zorgani, M. A., Quentin, R., and Lartigue, M. F. (2016). Regulatory RNAs in the less studied streptococcal species: from nomenclature to identification. *Frontiers in microbiology* 7, 1161.

4. ARTIGO 1: *IN SILICO* PREDICTION AND EXPRESSION PROFILE ANALYSIS OF SMALL NON-CODING RNAS IN *Herbaspirillum seropedicae* SmR1

Tatiane Dobrzanski[1], Vania Pobre[2], Leandro Moreno[1], Helba Cirino[3], Rose Adele Monteiro[1,3], Fábio de Oliveira Pedrosa[1], Emanuel Maltempi de Souza[1], Cecilia Maria Arraiano[2], Maria Berenice Reynaud Steffens[1,3]

[1]Department of Biochemistry and Molecular Biology, Universidade Federal do Paraná (UFPR), Av. Coronel. Francisco H. dos Santos, 81.531-980. Curitiba, Brazil.

[2]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157, Oeiras, Portugal

[3]Graduate Program in Bioinformatics, Universidade Federal do Paraná (UFPR), Rua Alcides Vieira Arcoverde, 1225; 81520-260, Curitiba, Brazil.

4.1. ABSTRACT

Herbaspirillum seropedicae is a diazotrophic bacterium from the β-proteobacteria class that colonizes endophytically important gramineous species, promotes their growth through phytohormone-dependent stimulation and can express *nif* genes and fix nitrogen inside plant tissues. Due to these properties this bacterium has great potential as a commercial inoculant for agriculture. The *H. seropedicae* Smr1 genome was completely sequenced and annotated but despite the availability of diverse structural and functional analyses of this genome, studies involving small non-coding RNAs (sRNAs) has not yet been done. We have conducted computational prediction and RNA-seq analysis to select and confirm the expression of sRNA genes in the *H. seropedicae* SmR1 genome, in the presence of two nitrogen independ sources and in presence of the flavonoid naringenin which simulates plant. This approach resulted in a set of 117 sRNAs distributed in riboswitch, cis-encoded and trans-encoded categories and among them 20 have Rfam homologs. Some sRNAs expression were confirmed *in vitro* and this work opens novel approaches to further understand the genetic regulation in this bacterium.

Keywords: *Herbaspirillum seropedicae* SmR1, sRNA, regulatory RNA, nitrogen fixation, diazotrophic bacterium, bacterial plant interaction

4.2. INTRODUCTION

Herbaspirillum seropedicae SmR1 is a diazotrophic and endophytic bacterium that belongs to the β-proteobacteria (Baldani *et al.*, 1996). This microorganism fixes nitrogen under microaerobic conditions inside the plant tissues of the economically important cereal crops wheat, rice, maize and sorghum (Roncato-Maccari *et al.*, 2003). *H. seropedicae* strains also appears associated with sugar cane and forage grasses (Pimentel *et al.*, 1991; Olivares *et al.*, 1996), fruit crops (Magalhaes Cruz *et al.*, 2001) and common bean (Schmidt *et al.*, 2011). Several studies have demonstrated the benefits of *Herbaspirillum*-plant interaction through the increase of the biomass of the inoculated plant (Gyaneshwar *et al.*, 2002; Baldani and Baldani, 2005; Estrada *et al.*, 2013; Alves *et al.*, 2014; Kandel *et al.*, 2017). Nitrogen cellular sources are the atmospheric dinitrogen and the nitrate, present in several environments. *H. seropedicae* SmR1 fixes nitrogen to ammonia in a reaction catalysed by the enzyme nitrogenase (Chubatsu *et al.*, 2012; Bonato *et al.*, 2016). The bacterial-plant interaction can promote plant growth and increase yield of crops since some compounds produced by the bacterium can stimulate the synthesis of phytohormones by plants (Bastián *et al.*, 1998). Plants also play an important role in establishing this interaction since they produce compounds that affect their associations with microorganisms. One of such compounds is naringenin, a flavonoid produced as secondary metabolite, that can stimulate or inhibit specific genetic responses in different bacteria associated with plants (Balachandar *et al.*, 2006; Hassan and Mathesius, 2012). It was demonstrated that naringenin stimulates the endophytic colonization of *Arabidopsis thaliana* by *H. seropedicae* Z67 (Gough *et al.*, 1997). In *H. seropedicae* SmR1, naringenin regulates the expression of several genes, positively or negatively (Tadra-Sfeir *et al.*, 2015). This microorganism can catabolize naringenin probably to obtain carbon and energy (Marin *et al.*, 2013).

The single circular chromosome of the *H. seropedicae* SmR1 strain was sequenced and 4,804 open reading frames were annotated (Pedrosa *et al.*, 2011). Recently 203 new genes were added by reannotation based on RNAseq data (unpublished). Since then, there are many studies focusing the genomic structure, gene expression and physiology of *H. seropedicae* SmR1 (Marin *et al.*, 2013; Tadra-Sfeir *et al.*, 2015; Balsanelli *et al.*, 2016; Bonato *et al.*,

2016; Pankievicz *et al.*, 2016; Batista *et al.*, 2018), but the investigation about the presence and function of small non-coding RNAs (sRNAs) was never performed. sRNAs have key regulatory roles in post-transcriptional control of gene expression. They can modulate turnover of target mRNAs and affect their translation (Waters and Storz, 2009; Lalaouna et. al. 2013, Carrier *et al.*, 2018). They can be found in all three domains of life and are particularly important in bacteria allowing them to rapidly respond to environmental challenges (Dennis and Omer, 2005; Zaratiegui *et al.*, 2007). These molecules are 50–500 nucleotides long and are located predominantly in intergenic or in untranslated regions in the bacterial genomes (Storz *et al.*, 2011). sRNAs can be divided in two major categories *trans*-encoded RNAs (traRNA) and *cis*-encoded RNAs (caRNA) (Zorgani *et al.*, 2016). The caRNA act at transcriptional level and are sensory RNAs elements such as riboswitches that adopt two conformational structures in response to chemical signals such as small ligands (Winkler and Breaker, 2005; Henkin, 2008; Narberhaus and Vogel, 2009). The traRNA comprises the trans-encoded sRNAs that are partially complementary to their target (Waters and Storz, 2009) and the antisense small RNAs (asRNAs) that are totally complementary to their target (Brantl, 2007; Saberi *et al.*, 2016). There are still traRNAs that bind to proteins to modulate their activity such as 6S RNA (Wassarman and Storz, 2000). The paring of many traRNAS to mRNA target sites is facilitated by the RNA chaperone Hfq, a Sm family protein, which binds to adenine- and uridine-rich sequences (AU-motif) in sRNA (Sauer, 2013; Updegrove *et al.*, 2016). *H. seropedicae* SmR1 contains a conserved Hfq protein with a classic hexameric ring shape, observed in all available Hfq structures, with sRNA and mRNA contact surfaces (Kadowaki et al, 2012).The presence of a variety of types of small non-coding RNAs provides a versatile regulation of metabolic functions (Gottesman and Storz, 2010; Bobrovskyy and Vanderpool, 2013; Dutta and Srivastava, 2018, Carrier *et al.*, 2018). The bioinformatic prediction of sRNAs followed by RNA-seq approach made possible genome screens for sRNAs and has shown that there are much more bacterial regulatory sRNAs than previously thought (Kahrstrom, 2012). In this study we applied *in silico* approach to predict sRNAs in *H. seropedicae* SmR1 genome and RNA-seq analysis to confirm their expression in bactéria gown in the presence of two nitrogen sources (ammonia or nitrate) and in the presence of naringenin. A set 117 of sRNAs transcripts were confirmed and some of them showed sequence identity with well-characterized sRNAs in other bacteria. sRNAs were experimentally detected confirming their existence.

4.3. MATERIALS AND METHODS

4.3.1. Bacterial Growth

H. seropedicae SmR1 (NCBI sequence: NC_014323.1) was grown at 30°C and with agitation of 120 rpm in NFbHPN medium containing 80 μg/mL streptomycin (Klassen *et al.*, 1997). Three growth conditions were used: (i) Control condition (CRT), bacteria grown on NFbHPN medium using malate as a source of carbon and 20 mmol/L NH_4Cl as nitrogen source; (ii) Naringenin condition (NAR), bacteria grown in NFbHPN medium in the presence of flavonoid naringenin (100μM); (iii) Nitrate condition (NIT), bacteria grown in NFbHP medium, using malate as carbon source and 10 mmol/L KNO3 as nitrogen source (Tadra-Sfeir et al., 2015; Bonato et al., 2016).

4.3.2. Screening of small RNAs by nocoRNAc and Cufflinks

The *H. seropedicae* SmR1 genome was screened with the computational tool nocoRNAc (Herbig and Nieselt, 2011) to search sRNAs by features that include promoter sequence, Rho-independent terminator and regions of sequence conservation. Furthermore, regions with high level of transcription, free of encoded proteins, were assessed in three individual RNA-seq data sets using *Cufflinks* (Trapnell *et al.*, 2012) to verify the presence/expression of sRNAs. The nocoRNAc (non-coding RNA characterization) tool predicts putative sRNAs based on analyses not limited to intergenic regions. In order to find the location of candidates, firstly the SIDD sites, which are destabilized regions in the genomic DNA, were identified as putative promoter regions. The SIDD calculation was conducted using default values. After that, the putative Rho-independent terminators were predicted by the TranstermHP program, which is integrated in nocoRNAc. The tool was run using the standard protocol, with the option overwrite set up as described in the user guide (Herbig and Nieselt, 2011). The coordinates of SIDD sites with Rho-independent terminator are used for generating a list of putative sRNAs. Thus, the nocoRNAc tools may detect putative sRNAs in whole genome, even those which are encoded antisense from protein genes.

We assessed RNA-seq data to uncover sRNAs based only on read alignments. The reads from RNA-seq were mapped to *H. seropedicae* SmR1 genome with the tRNA, mRNA,

rRNA and their 50 flaking nucleotides masked. We used the default parameters on the Cufflinks program to localize transcribed/expressed regions on the genome likely to encode sRNAs.

4.3.3. Mapping and visualization of sequence reads and analyses of predicted sRNAs

Before mapping the short reads to the *H. seropedicae* SmR1 reference genome, the rRNA sequences were masked using the cross-match program. Recursive trimming of the reads at 5' and 3' to 35 nucleotides were performed using a Perl script and the Mate-Paired reads were aligned to the reference genome using the alignment tool Short Read Mapping Program SHRiMP (Rumble *et al.*, 2009). The program was set up to tolerate 3 mismatches. The maximal number of hits to each read was 1. Samtools (Li *et al.*, 2009) was used to convert data into SAM/BAM format. Mapped RNA-seq reads in BAM format were visualized in the genome browse Artemis (Carver *et al.*, 2012). A sequence predicted as a sRNA was considered expressed when the minimum read coverage was 5-fold.

4.3.4. RNA -Seq data analysis

The RNA-seq data sets used in this work are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-3435 e E-MTAB-2842. Small non-coding RNA expression profiles was obtained with Artemis (Rutherford *et al.*, 2000). We employed an RPKM normalized expression values of three RNA-seq conditions describe in Tadra-Sfeir *et al.* 2015 and Bonato *et al.* 2016 for the heatmap and hierarchical clustering. We used Pearson correlation as distance measure and average linkage as clustering method in Heatmapper (Babicki *et al.*, 2016). Genome coordinate plot of non-coding RNA was performed with DNAPlotter (Carver *et al.*, 2009).

4.3.5. RNA extraction and northern blot analysis

Overnight cultures grown in the CRT, NAR and NIT conditions were diluted in fresh medium to an initial OD_{600}= 0.1 and grown to exponential (OD_{600} 0.7) and stationary phase (10h of growth). Culture samples were withdrawn and mixed with an equal volume of RNA stop buffer (10 mM Tris at pH 7.2, 5 mM $MgCl_2$, 25 mM NaN_3, and 500 µg/mL

chloramphenicol). The total RNA extraction followed the protocol of cell lysis and phenol:chloroform extraction (adapted from (Andrade *et al.*, 2012)). After a precipitation step in ethanol and 300 mM sodium acetate, RNA was resuspended in MilliQ-water. The integrity of RNA samples was evaluated by agarose gel electrophoresis. When necessary, Turbo DNase (Ambion) treatment following a new phenol: chloroform step was used to remove contaminant DNA. Next, 10– 20 μg of total RNA was used to analyse small RNA expression on 10% polyacrylamide gels in TBE 1x buffer. RNA was transferred onto Hybond-N+ membrane (Amersham Biosciences) using TAE 1x as transfer buffer. RNAs were UV cross-linked to the membrane with a UVC 500 apparatus for 3 min (Amersham Biosciences). DNA templates carrying a T7 promoter sequence for *in vitro* transcription were generated by PCR using genomic DNA of *H. seropedicae* SmR1 and the primers listed in Table 1. Hsnc028 and Hsnc082 were detected by 5' -end labelling of an antisense primer (Table1). Radiolabelled probes using rUTP α-^{32}P (T7 probes) or γ-^{32}P ATP (primer probes) were purified on G25 Microspin columns (GE Healthcare). Hybridizations were carried out overnight at 42°C or 68°C with the PerfectHyb Plus Hybridization Buffer (Sigma). RNA Decade marker (Ambion) was used when detecting non-coding RNAs up to 150 nt; for longer transcripts, the 100–1000 bp Ladder (Biotools) was used. All radiochemicals were purchased from Perkin-Elmer.

TABLE 1. Oligonucleotides used in radiolabelling reactions

Probe	Sequence (5' – 3')
ssrS-F	CCGTGTTCGCGATTGCC
ssrS-T7	<u>TAATACGACTCACTATAGG</u>CCGGCATCCTGAACCTG
Hsnc042-F	GATGCCCGACTGCTGAAACG
Hsnc042-T7	<u>TAATACGACTCACTATAGG</u>TAGCGTCGGAATCGCGTTCCTG
Hsnc073-F	GCAATAACCAATGCGCAGG
Hsnc073-T7	<u>TAATACGACTCACTATAGG</u>GCATCATCAAGGGATGCCAG
Hsnc028	AAATCAGGCGTTTGTCATGGTTCGGTAAG
Hsnc082	AACGATGGAAGTACGGTGGTTCGCGTGATG

The T7 promoter sequence in the oligos is underlined.

4.4. RESULTS

4.4.1. sRNAs in the *H. seropedicae* SmR1 genome

To search sRNAs in the genome of *H. seropedicae* SmR1 we used the nocoRNAc software, a bioinformatic tool that predicts sRNAs based on the co-localization of transcriptional terminators and promoter and is not limited to intergenic regions (Herbig and Nieselt, 2011). We identified 769 putative sRNAs. At the same time, we verified the presence of sRNAs transcripts in the RNA-seq data of *H. seropedicae* SmR1 using Cufflinks. We were able to identify 1,395 regions being transcribed which could encode sRNAs. We analysed three RNA-seq data conditions obtained during the exponential growth phase of the bacterium: (i) control (CRT) - bacteria grown in NFbHPN medium containing NH_4Cl as nitrogen source, (ii) presence of naringenin (NAR) - bacteria cultured in NFBPN medium containing NH_4Cl and in the presence of flavonoid naringenin, and (iii) nitrate (NIT) - bacteria grown in NFBP medium containing KNO_3 as nitrogen source. Using the coverage criterion which established a minimum coverage ≥5 as a confidence level to select sRNAs in at least one of the three culture conditions we were able to confirm 117 sRNAs transcripts in *H. seropedicae* SmR1 which have been termed Hsnc001 to Hsnc117 (Additional data 1). Forty sRNAs (34.5%) resulted only from Cufflinks, 63 sRNAs (54.3%) resulted only from nocoRNAc and 14 (12.1%) appeared in both approaches. These sRNAs range in length from 41 to 560 nucleotides and are equally distributed in the genome of *H. seropedicae* SmR1 being 67 annotated on the sense and 50 on the antisense strand in intergenic regions (Figure 1A). According to the genomic location the sRNAs were distributed in riboswitch (10), cis-encoded (26) and trans-encoded (81) categories (Figure 1B). Regarding base composition sRNAs molecules have an average GC content of 54.97% whereas the genome has about 63.4% GC (Additional data 1). We observed that the riboswitches present high GC content (61.8%) and the known housekeeping sRNAs tmRNA, ssrS and 4.5S had 52.3, 53.7 and 65.7% GC content, respectively.

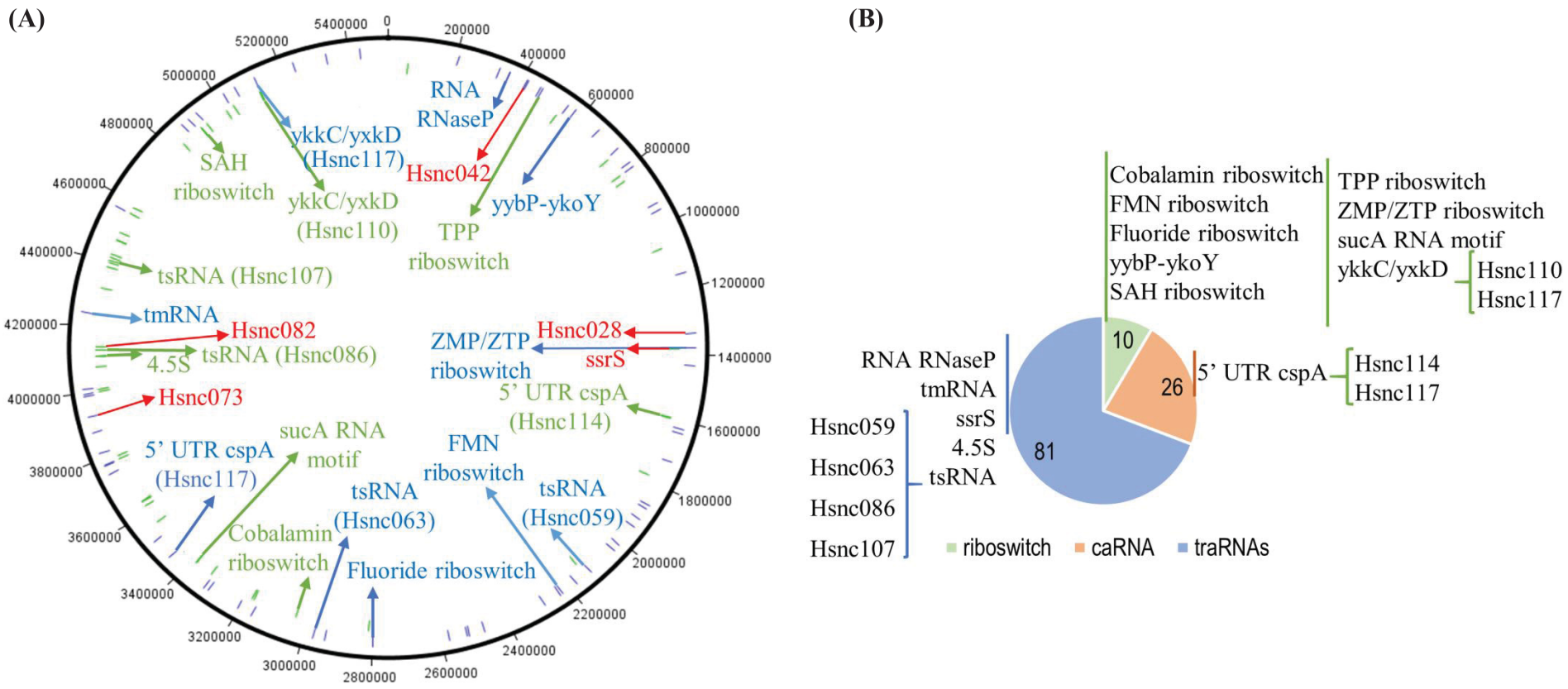


FIGURE 1. Identification of the sRNAs predicted by bioinformatics of *H. seropedicae* SmR1. (A) Distribution of 117 sRNAs in the genome of *H. seropedicae* SmR1. The sRNAs annotated on the sense DNA strand are marked in blue and the annotated on the antisense strand in green. The sRNAs with defined identity (homology given by Rfam) are indicated in green or blue and the selected for experimental validation are indicated in red. **(B)** Categorization of predicted sRNA of *H. seropedicae* SmR1. The programs identified cis-encoded, trans-encoded and riboswitch sRNAs. The sRNAs with defined identity are indicated in each group.

4.4.2. Identity of sRNAs of *H. seropedicae* SmR1

As the sRNAs show conserved key nucleotide sequences among bacterial and archaeal species we sought *H. seropedicae* SmR1 sRNAs with sequence identity with other sRNAs already known comparing the sRNA sequences in the RNA database (Rfam). Only 20 sRNAs had sequence identity with well-characterized sRNAs in other bacteria (Table 2). This result suggests that most of the *H. seropedicae* SmR1 sRNAs may be new or present low level of identity with those deposited in the Rfam database. Among the sRNAs identified the Toxic small RNA (tsRNA) and sucA RNA motif were found essentially in β-proteobacteria (Sass, 2015; Greenlee *et al.*, 2018) whereas YkkC/YxkD leader is present in some Cyanobacteria and Proteobacteria (Barrick *et al.*, 2004). The sRNAs belonging to the family of small toxic RNAs in *H. seropedicae* SmR1 were Hsnc59, Hsnc63, Hsnc86 and Hsnc107 (Table 2). These small Toxic RNAs were found to be expressed in several strains of *Burkholderia cenocepacia* and, although they do not present a known function, they are capable of inhibiting *Escherichia coli* growth when introduced in a cloning vector (Kimelman *et al.*, 2012; Sass, 2015). The Hsnc006 was annotated as the sucA 5'UTR which is considered a riboswitch candidate since the ligand that changes its conformation is still unknow (Sun *et al.*, 2013). This sequence is the 5' UTR of *sucAsucBlpd* operon and, according to RNA-seq data, the RNA motif and the operon exhibited proportional expression level in the three conditions analysed (Additional data 1). We found two copies of YkkC/YxkD leader (Hsnc110 and Hsnc116) upstream HSERO_RS22365 and HSERO_RS22370, respectively, which encode two lipid kinases involved in the inorganic ion transport and metabolism. Recently this riboswitch was renamed as guanidin-I riboswitch since it senses and responds to guanidine and controls genes that modify or pump guanidine as a toxic compound of bacteria (Nelson *et al.*, 2017).

TABLE 2. *H. seropedicae* SmR1 sRNAs identified in Rfam

Predicted			Rfam					
sRNA	size	GC%	ID	Acession	Start	End	Bit score	e-value
Hsnc001	99	65.66	4.5S	RF00169	1	99	76.4	8E-19
Hsnc002	111	60.36	Fluoride riboswitch	RF01734	8	81	50.3	4.4e-10
Hsnc006	94	56.38	sucA RNA motif	RF01070	12	93	79.8	2.3e-16
Hsnc029	100	61.00	ZMP/ZTP riboswitch	RF01750	1	100	59.3	2.6e-09
Hsnc035	169	65.09	FMN riboswitch	RF00050	1	169	112.6	5.3e-28
Hsnc050	177	53.67	ssrS (6S)	RF00013	1	177	67.3	2.5e-14
Hsnc059	93	49.46	Betaproteobacteria toxic RNA	RF02278	25	93	59.6	1.4e-12
Hsnc063	115	43.48	Betaproteobacteria toxic RNA	RF02278	48	115	62.4	3.4e-13
Hsnc083	384	52.34	tmRNA	RF00023	1	381	193.0	4.6e-57
Hsnc086	97	47.42	Betaproteobacteria toxic sRNA	RF02278	34	97	64.7	7.6e-14
Hsnc107	98	43.88	Betaproteobacteria toxic sRNA	RF02278	31	97	61.1	6.1e-13
Hsnc109	293	58.36	TPP riboswitch	RF00059	86	196	55.6	2.3e-10
Hsnc110	101	64.36	ykkC-yxkD	RF00442	1	101	99.0	1.8e-23
Hsnc111	335	61.19	RNA RNaseP	RF00010	1	335	212.1	1.1e-68
Hsnc112	184	63.04	yybP-ykoY	RF00080	16	184	47.7	6E-12
Hsnc113	90	65.56	SAH riboswitch	RF01057	1	90	49.1	9.6e-09
Hsnc114	373	51.47	5' UTR cspA	RF01766	1	373	94.3	1.6e-24
Hsnc115	247	63.97	Cobalamin riboswitch	RF00174	1	247	111.2	1.6e-30
Hsnc116	100	60.0	ykkC-yxkD	RF00442	1	100	93.4	4.8e-22
Hsnc117	387	48.97	5' UTR cspA	RF01766	1	388	86.7	2.7e-22

We also found the TPP (Hsnc109) and FMN (Hsnc035) riboswitches. The TPP riboswitch is immediately upstream of HSERO_RS02120 (*thiC*) encoding the phosphomethylpyrimidine synthase and is known to bind directly to thiamine pyrophosphate (TPP) turning off TPP biosynthesis (Mehdizadeh Aghdam *et al.*, 2017). The FMN riboswitch is upstream of HSERO_RS09820 (*ribE*) encoding the 6,7-dimethyl-8-ribitylllumazine synthase which catalyses one of last steps in the biosynthesis of riboflavin. FMN binds to the FMN aptamer and regulates the *ribE* expression (Pedrolli *et al.*, 2015).

H. seropedicae SmR1 still presents the SAH riboswitch (Hsnc113), Cobalamin riboswitch (Hsnc115), ZMP / ZTP riboswitch (Hsnc029), Yybp-ykoY (Hsnc112) and Fluoride riboswitch (Hsnc002). The SAH riboswitch is upstream HSERO_RS21435, encoding S-adenosyl- (L) -homocysteine (SAH), and is involved in S-adenosyl- (L) - methionine (SAM) regeneration cycle (Wang *et al.*, 2008; Edwards *et al.*, 2010). Cobalamin riboswitch is upstream HSERO_RS13325-HSERO_RS13320 operon encoding the cobalt transporter CbtB-CbtA acting in concert with vitamin B12 biosynthesis systems (Rodionov *et al.*, 2003; Polaski *et al.*, 2017). ZMP/ZTP riboswitch regulates the expression of carbon metabolism genes (Ducker and Rabinowitz, 2015; Kim *et al.*, 2015). *H. seropedicae* SmR1 showed this riboswitch (Hsnc029) upstream to the HSERO_RS06140 (*glyA*) encoding serine hydroxymethyltransferase, a pyridoxal phosphate-dependent enzyme that plays an important role in the cellular pathways of a carbon. Yybp-ykoY is a manganese riboswitch that binds directly to Mn^{2+} and is associated with YebN/MntP genes (Dambach *et al.*, 2015; Price *et al.*, 2015; Li *et al.*, 2018). In *H. seropedicae* SmR1, this riboswitch is located upstream HSERO_RS02630 encoding the manganese efflux pump MntP. In *Xanthomonas oryzae* this riboswitch acts as an essential Mn^{2+} sensor in infections during interaction with rice (Li *et al.*, 2018). Fluoride riboswitch located upstream of HSERO_RS12335 coding the voltage-gated chloride channel protein. This riboswitch has been experimentally verified by (Liu *et al.*, 2017) detecting fluoride and triggering the expression of genes that can help *Enterobacter cloacae* FRM to mitigate fluoride toxicity, using a fluorine carrier to expel fluoride from the cells.

We found only two sRNAs with identity in the cis-encoded category, the *cspA* 5'UTR mRNAs Hsnc114 and Hsnc117. These sRNAs were identified in *H. seropedicae* SmR1 in untranslated regions of genes HSERO_RS07020 and HSERO_RS15195 encoding cold-shock proteins. These elements are known to be involved in the expression of gene *cspA* in response to temperature shift (Jiang *et al.*, 1996; Zhang *et al.*, 2018), however it was already demonstrated that they might have a role in stress tolerance (Derman *et al.*, 2015). Both *H. seropedicae* cspA 5 'UTR mRNAs contain single-strand AU-motif in mRNA which could be binding sites of the RNA chaperone Hfq, as demonstrated in *E. coli* (Hankins *et al.*, 2010).

Some well-preserved housekeeping sRNAs were also found in *H. seropedicae* SmR1, ssrS or 6S RNA (Hsnc050), 4.5S RNA (Hsnc001) and tmRNA (Hsnc083). These sRNAs associate with proteins and are highly expressed in the cell. The 6S interacts with the primary form of holoenzyme of RNA polymerase, negatively regulates transcription and is involved in

modulating stress and optimizing survival during nutrient limitation (Steuten *et al.*, 2014). The 4.5S is part of the signal recognition particle (SRP) ribonucleoprotein complex (Mercier *et al.*, 2017). In most bacteria the SRP consists of an RNA molecule (4.5S) and the Ffh protein that bind to ribosome stopping protein synthesis. The tmRNA (transfer messenger RNA) forms a ribonucleoprotein complex (tmRNP) that binds to bacterial ribosomes that are blocked in the middle of protein synthesis; it is able to recycle the blocked ribosome by bringing a stop codon and adding a proteolysis-inducing tag to unfinished polypeptides (Sundermeier *et al.*, 2005). Although the majority of the *H. seropedicae* SmR1 sRNAs presented in this work do not have sequence homology, some sRNAs probably show function homology.

4.4.3. Expression of sRNA in *H. seropedicae* SmR1

Bacteria have a versatile system to respond quickly to environmental changes. In this process, many sRNAs are often expressed to regulate gene expression in a specific or different conditions and stages of growth. Thus, we wanted to determine how sRNAs are expressed in different culture conditions. The reading count of exponential phase in RNA-seq data was performed for each sRNA followed by normalization by Reads Per Kilobase Million (RPKM). A heat map that includes the 117 sRNA reveals different expression profiles according to the growth condition of *H. seropedicae* SmR1 (Figure 2). We observed that a large cluster with 62 sRNAs are more expressed in the nitrate condition than in the control and naringenin conditions. Another cluster with 24 sRNAs is more expressed in the naringenin condition than in the control and nitrate conditions. This demonstrates the role of environment in the expression of sRNAs and the influence of sRNAs when the bacterium is exposed to certain nutritional conditions.

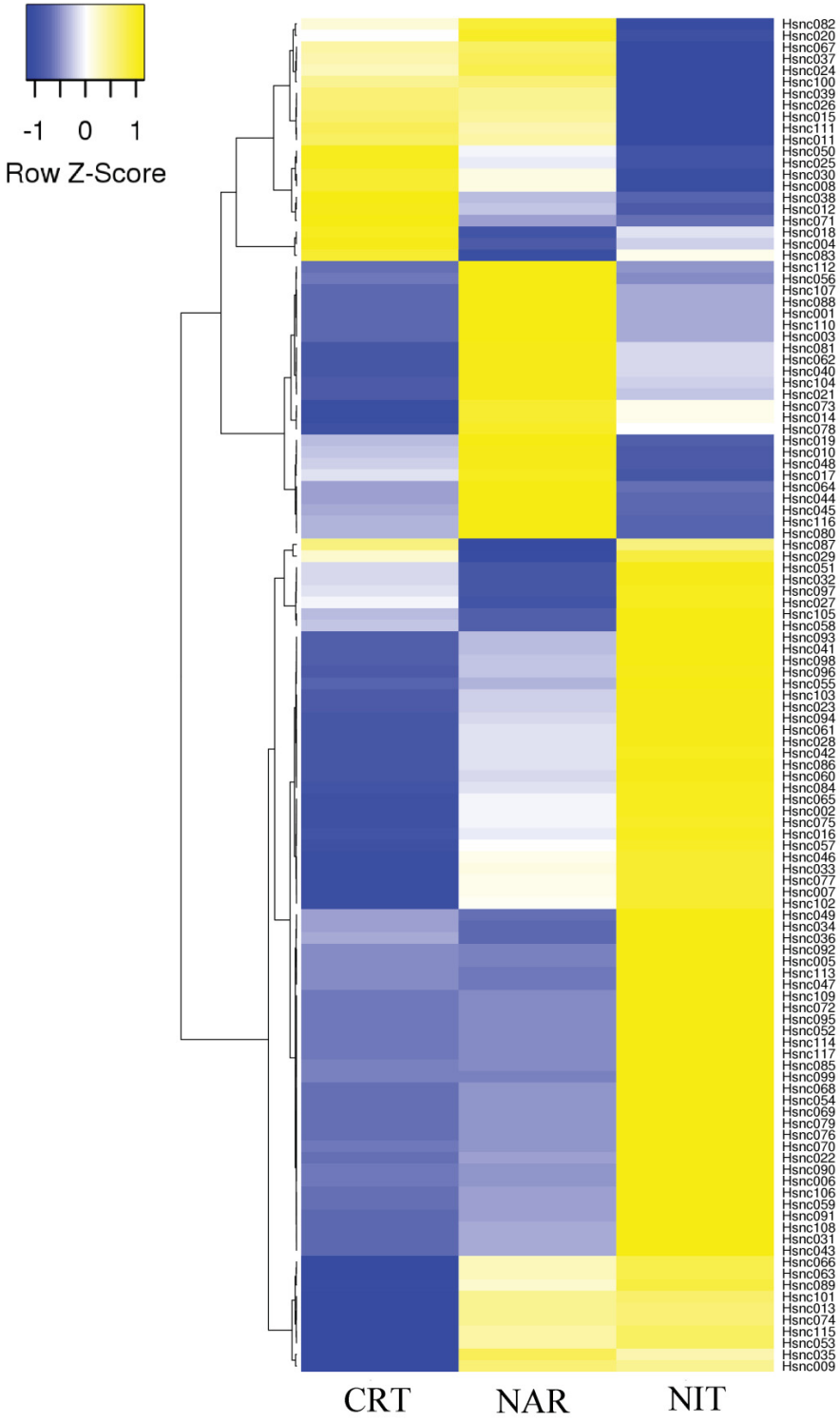


FIGURE 2. Heatmap showing relative expression levels of 117 sRNAs in the genome of *H. seropedicae* SmR1. The heatmap showed the expression levels of sRNAs in the control culture condition (CRT), presence of naringenin (NAR) and nitrate (NIT) during the exponential phase. The dendrogram provides the visualization of a hierarchical clustering of sRNAs with similar expression patterns. A scale of z-score relation to colour intensity is shown.

4.4.4. Experimental validation of *H. seropedicae* SmR1 sRNAs

We also used the coverage criterion of RNA-seq reads greater than or equal to five, in at least one of the three culture conditions (Additional data 1), to validate their expression experimentally by northern blot. Although coverage values do not represent expression quantification (as opposed to RPKM quantification), they are an indicator that a given sRNA may be expressed. We choose the ssrS (Hsnc050) as a control, since it is a conserved housekeeping sRNA among bacterial species (Barrick *et al.*, 2005; Trotochaud and Wassarman, 2005; Steuten *et al.*, 2014).

Considering that many sRNAs are induced under stress conditions, such as the lack of nutrients in the stationary phase, we evaluate the expression of the sRNAs in two phases of growth, exponential (DO_{600} = 0.7) and stationary (after 10 hours of culture). *H. seropedicae* SmR1 was cultured under the CRT, NAR and NIT conditions and the total RNA was extracted and hybridized with specific radiolabelled probes. We were able to confirm that the five selected sRNAs are expressed in *H. seropedicae* SmR1 (Figure 3). It is notable that all RNAs were expressed in all analysed growth conditions (CRT, NAR and NIT) as well as in the exponential and stationary phases (Figure 3).

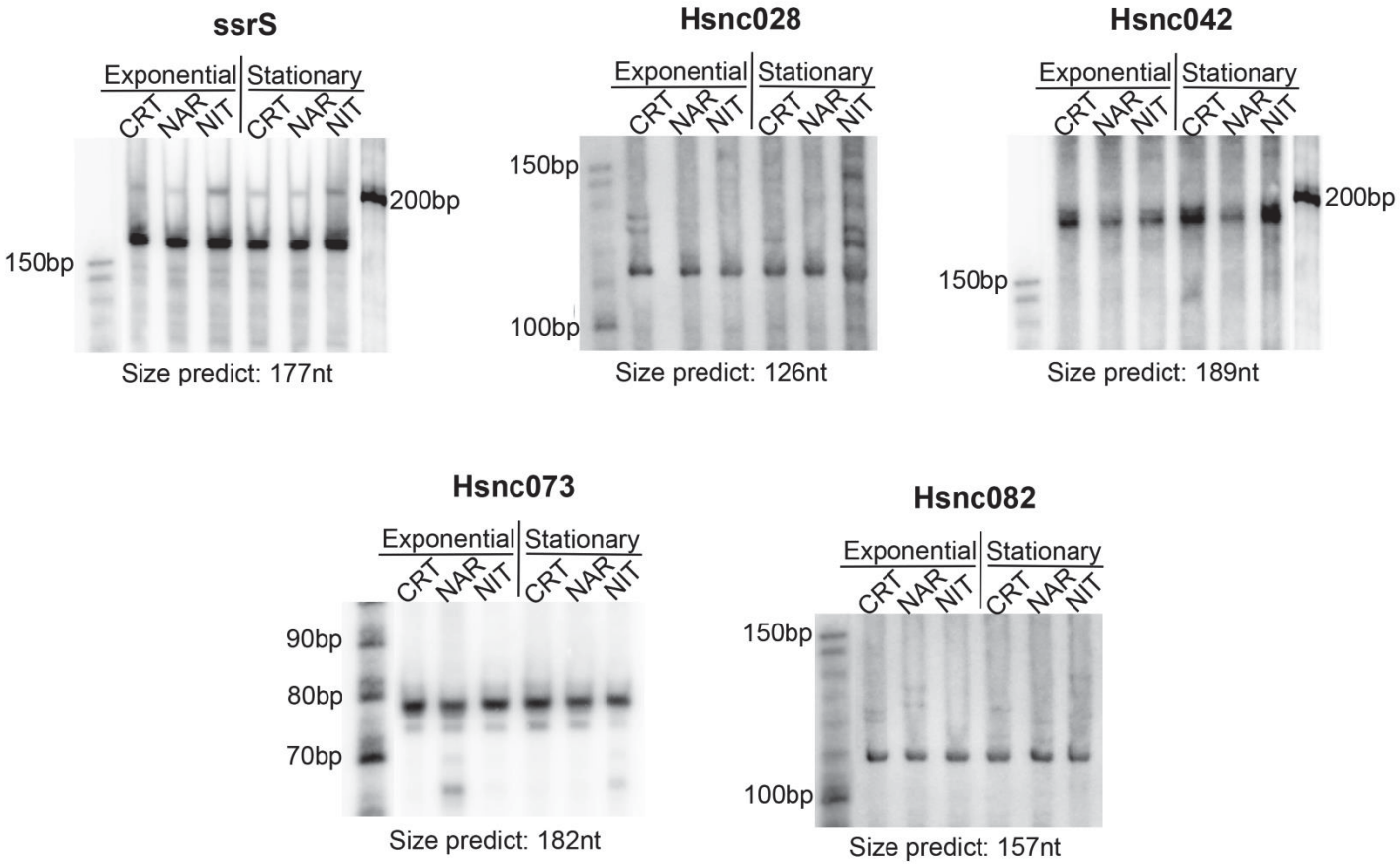


FIGURE 3. Validation of ssrS, Hsnc028, Hsnc042, Hsnc073 and Hsnc082 sRNAs of *H. seropedicae* SmR1 by northern blot. The expression of sRNAs was analysed in the exponential phase (OD_{600} ~ 0.7) and in the stationary phase (after 10h growth), in the control condition (CRT), in the presence of naringenin (NAR) and in the nitrate (NIT) condition. Markers are indicated to the right and left in the images.

The function of these sRNAs is yet to be elucidated. In addition, we notice that Hsnc073 and Hsnc082 sRNAs showed lengths in the northern blot smaller than length initially predicted by bioinformatics (Figures 3 and 4). We suggest the smaller lengths of Hsnc073 and Hsnc082 sRNAs may be due to sRNA processing since most of the sRNAs are transcribed with larger length and then later processed by RNases for smaller functional lengths (Storz *et al.*, 2011). This difference can be also due to an imprecision in the prediction of sRNAs by the bioinformatics tools used in this work. Previously, we already observed that the nocoRNAc tool can predict sRNAs with larger than expected lengths, as the case of ssrS (6S). This sRNA was initially predicted with a length of 327 nucleotides by nocoRNAc, however based on RNA-seq we corrected its length to 177 nucleotides and, in fact, an RNA band with around this length was obtained in the northern blot (Figure 3). Bacterial ssrS (6S) RNAs are generally transcribed as pre-6S RNA and then processed in 5' end by ribonucleases that cuts a short sequence to mature form (Burenina *et al.*, 2015). The *H. seropedicae* 6S RNA length is very close to the 6S of *Neisseria meningitidis* MC58 and *Pseudomonas aeruginosa* which are about 180 nucleotides (Barrick *et al.*, 2005). The Hsnc028 (126 nucleotides) and Hsnc073 (182 nucleotides) also presented well distributed coverage in the RNA-seq profile (data not shown).

4.5. DISCUSSION

In this work we identified and validated sRNAs in β-proteobacteria *H. seropedicae* SmR1, a highly versatile diazotrophic bacterium capable of metabolizing a wide range of carbon and nitrogen sources. Initially, we showed that 2164 candidate sRNAs were predicted by bioinformatics tools Cufflinks and nocoRNAc that uses a specific prediction method for prokaryotic sRNA. To confirm which of these sRNAs are expressed in *H. seropedicae* SmR1, we analysed RNA-seq data from three culture conditions (CRT, NAR and NIT) and confirmed that *H. seropedicae* SmR1 expresses about 117 sRNAs, each expressed in at least one of the three conditions analysed. The number of sRNAs predicted by bioinformatics tools varies among species *Streptomyces coelicolor* presented about 843 sRNAs predicted by RNAz and nocoRNAc (Herbig and Nieselt, 2011). *Burkholderia pseudomallei*, a Betaproteobacterium, presented about 1306 sRNAs predicted by bioinformatics (Khoo *et al.*, 2012) whereas in *B. cenocepacia* J2315 were predicted about 213 sRNAs (Coenye *et al.*, 2007).

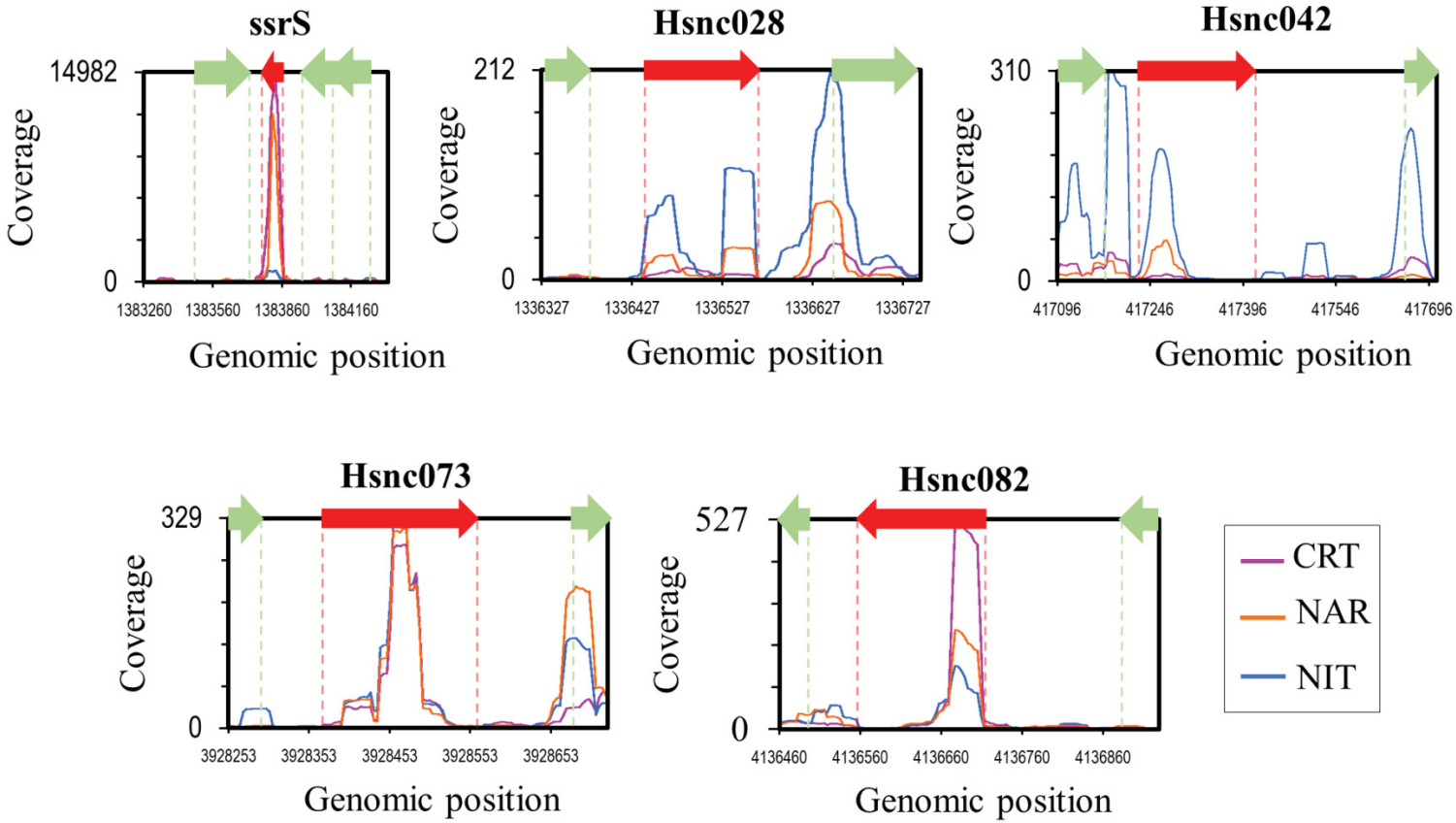


FIGURE 4. Coverage profiles of *H. seropedicae* SmR1 sRNAs detected by RNA-seq. The y-axes indicate the reading coverage for each sRNA in the control conditions (purple lines), in the presence of naringenin (orange lines) and nitrate (blue lines). The x-axes denote the genomic positions according to the *H. seropedicae* SmR1 genome coordinates.

When we compared the sequences of the *H. seropedicae* SmR1sRNAs with the RNA family database (Rfam), we found only 20 sRNAs with sequence identity similar to some RNA family already known. We were able to identify about 26 cis-encoded sRNAs in *H. seropedicae* and SmR1 being two of them belonging to the family of *cspA* sRNAs according to Rfam. The 10 riboswitches of this organism showed definite sequence identity. Of the 81 trans-encoded sRNAs only 8 showed identity given by Rfam. The occurrence of a large number of unidentified *H. seropedicae* SmR1 sRNAs are not surprising since most of the sRNAs identified in many other organisms still do not have known function. In addition, sRNAs are most characterized in model organisms such as *E. coli*, *Salmonella enterica* and *Staphylococcus aureus* (Sharma and Heidrich, 2012) that are very distant from *H. seropedicae*. Thus, we suggest that *H. seropedicae* SmR1 may exhibit specific sRNAs or a degree of identity closer to other organisms belonging to its same class. In β-proteobacteria the sRNAs have been described in *Burkholderia* species although they were not well functionally characterized (Sass, 2015; Ghosh *et al.*, 2017; Sass *et al.*, 2017).

Many processes could be controlled by bacterial sRNAs induced by specific metabolic or environmental signals (Michaux *et al.*, 2014) or expressed constitutively under different growth conditions, such as housekeeping sRNAs (6S, tmRNA and 4.5S) that participate in the regulation of genes that are constitutively expressed (Hershberg *et al.*, 2003). *H. seropedicae* SmR1 may probably express a larger number of sRNAs whose expression is induced in other culture metabolic conditions than those observed in the three RNA-seq assays.

Interestingly, all *H. seropedicae* SmR1 sRNAs had a mean GC content (54.97% GC) below the mean of the genome content (63.4% GC). However, there are some sRNAs with slightly higher GC content such as riboswitches (61.81% GC) and known housekeeping sRNAs (6S, tmRNA and 4.5S with an average 57.22% GC). RNAs that exhibit high GC content generally exhibit a more rigid and conserved structure. This conserved structure is necessary for the regulation of specific target molecules, such as maintenance genes or molecules that hardly evolve. In contrast, sRNAs that have a more flexible structure with a low GC content are probably involved in regulating the expression of several different genes or molecules that frequently evolve (Hershberg *et al.*, 2003, Zhao *et al.*, 2010, Sass *et al.*, 2017). We have observed that the four Toxic sRNAs have low GC content (Hsnc063 43.48%, Hsnc086 47.42%, Hsnc107 43.88% and Hsnc059 49.46%). The Toxic sRNAs currently found exclusively in β-proteobacteria are trans-encoded and Kimelman *et al.* (2008) proposed that these sRNAs interact with several genes that interfere in the ribosomal binding site (RBS).

We suggest that these sRNAs have a more flexible structure with low GC content to control the regulation of different targets.

We created a heatmap to observe the expression of sRNAs between control and different culture conditions and observed a large cluster of sRNAs more expressed in the NIT condition. The main source of nitrogen for most organisms is ammonium and, in the absence or low concentration of ammonium, bacteria need to mobilize alternative nitrogen sources to maintain growth and increase chances of survival. In the absence or restriction of ammonium, *H. seropedicae* SmR1 can assimilate nitrate (Bonato *et al.*, 2016). The RNA-seq profile revealed that the change in nitrogen source from ammonium to nitrate caused modifications in the pattern of gene expression in *H. seropedicae* SmR1, more than 37% of the genes were differentially expressed in the nitrate condition and the carbon consumption was increased (Bonato *et al.*, 2016). Since sRNAs may play an important role in nutritional deprivation (Michaux *et al.*, 2014), our data suggest that when the nitrate is the only nitrogen source many sRNAs may be influencing the post-transcriptional regulation of genes involved in carbon and nitrogen metabolism.

A small cluster of sRNAs more expressed in the presence of naringenin was also observed in the heatmap. Naringenin is a plant-derived flavonoid and may act as a signal molecule during endophytic colonization of *H. seropedicae* SmR1 (Tadra-Sfeir *et al.*, 2011). In this bacterium, naringenin triggers a change in gene expression to reduce motility and flagella synthesis (Tadra-Sfeir *et al.*, 2015). An extensive bacterial sensory system for adaptation and survival to the plant environment was also observed during the early stages of colonization of maize (Balsanelli *et al.*, 2016). Our data suggest that many sRNAs may be involved in the post-transcriptional regulation of genes related to the adaptation and endophytic colonization of *H. seropedicae* SmR1 in plants. Further investigation is required to determine if some sRNAs may be related to flagella synthesis and bacterial motility as shown in other organisms (Bak *et al.*, 2015; Fuentes *et al.*, 2015).

We were able to validate by northern blot the expression of the Hsnc050 (6S RNA), Hsnc028, Hsnc042, Hsnc073 and Hsnc082 sRNAs. Hsnc073 and Hsnc082 sRNAs had smaller than expected lengths according to the northern blot. As we said earlier this difference could be due to sRNA processing or even imprecision in the prediction. (Klein *et al.*, 2002) also observed a length difference of sRNAs predicted for *Pyrococcus furiosus*, of the 11 sRNAs identified in Northern blot only one showed observed length corresponding to the initially predicted size. *Sinorhizobium meliloti* also showed sRNAs with smaller lengths than

predicted for sRNAs 6S and sra25 (15nt of difference) as well as for sm84 (30nt of difference) and sm270 (10nt of difference) (Valverde *et al.*, 2008). The sRNA processing requires enzymatic cleavage to remove extra residues by ribonucleases to generate functional stable forms (Matos *et al.*, 2017). *H. seropedicae* SmR1 presents RNase E (Hsero_RS09410) an endoribonuclease which preferentially cleaves AU-rich regions (Arraiano *et al.*, 2010) and affect sRNA biogenesis (Chao *et al.*, 2017). PNPase (Hsero_RS08755) could also trim sRNAs contributing to their maturation and/or degradation (Andrade *et al.*, 2012). In fact, this difference in size can also account for the differences in expression observed between the RNA-Seq data and the Northern blot.

4.6. CONCLUSIONS

We reported the expression of several sRNA in *H. seropedicae* SmR1 genome in the presence of two nitrogen sources and/or in the presence of naringenin. The functions of the novel sRNAs remain unknown but their existence in this bacterium confirms the evidence that sRNAs are involved in many different cellular activities to adapt to nutritional and environmental changes. Some of them may participate in the regulation of nitrogen metabolism or in the bacterial-plant interaction. The discovery and knowledge of these sRNA molecules in this nitrogen fixation bacterium is very important due to its biotechnological significance.

ACKNOWLEDGMENTS

We are grateful to Michelle Zibettti Tadra-Sfeir and Paloma Bonato for making RNA-seq data available. Roseli A. Prado, Marilza D. Lamour, and Valter A. de Baura are acknowledged for technical assistance to the Nitrogen Fixation Group.

4.7. REFERENCES

Alves, G.C., Videira, S.S., Urquiaga, S., and Reis, V.M. (2014). Differential plant growth promotion and nitrogen fixation in two genotypes of maize by several *Herbaspirillum* inoculants. *Plant and Soil* 387(1-2)**,** 307--321. doi: 10.1007/s11104-014-2295-2.

Andrade, J.M., Pobre, V., Matos, A.M., and Arraiano, C.M. (2012). The crucial role of PNPase in the degradation of small RNAs that are not associated with Hfq. *RNA* 18(4)**,** 844-855. doi: 10.1261/rna.029413.111.

Arraiano, C.M., Andrade, J.M., Domingues, S., Guinote, I.B., Malecki, M., Matos, R.G., *et al.* (2010). The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol Rev* 34(5)**,** 883-923. doi: FMR242 [pii]10.1111/j.1574-6976.2010.00242.x.

Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., *et al.* (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic acids research* 44(W1)**,** W147--W153. doi: 10.1093/nar/gkw419.

Bak, G., Lee, J., Suk, S., Kim, D., Young Lee, J., Kim, K.S., *et al.* (2015). Identification of novel sRNAs involved in biofilm formation, motility, and fimbriae formation in *Escherichia coli. Sci Rep* 5**,** 15287. doi: 10.1038/srep15287.

Balachandar, D., Sandhiya, G.S., Sugitha, T.C.K., and Kumar, K. (2006). Flavonoids and growth hormones influence endophytic colonization and in planta nitrogen fixation by a diazotrophic Serratia sp. in rice. *World Journal of Microbiology and Biotechnology* 22(7)**,** 707--712. doi: 10.1007/s11274-005-9094-0.

Baldani, J.I., and Baldani, V.L.D. (2005). History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. *Anais da Academia Brasileira de Ciências* 77**,** 549-579.

Baldani, J.I., Pot, B., Kirchhof, G., Falsen, E., Baldani, V.L., Olivares, F.L., *et al.* (1996). Emended description of *Herbaspirillum*; inclusion of [*Pseudomonas*] *rubrisubalbicans*, a milk plant pathogen, as *Herbaspirillum rubrisubalbicans* comb. nov.; and classification of a group of clinical isolates (EF group 1) as *Herbaspirillum* species 3. *Int J Syst Bacteriol* 46(3)**,** 802-810. doi: 10.1099/00207713-46-3-802.

Balsanelli, E., Tadra-Sfeir, M.Z., Faoro, H., Pankievicz, V.C., de Baura, V.A., Pedrosa, F.O., *et al.* (2016). Molecular adaptations of *Herbaspirillum seropedicae* during

colonization of the maize rhizosphere. *Environmental microbiology* 18(8), 2343--2356. doi: 10.1111/1462-2920.12887.

Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., *et al.* (2004). New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proceedings of the National Academy of Sciences* 101(17), 6421--6426. doi: 10.1073/pnas.0308014101.

Barrick, J.E., Sudarsan, N., Weinberg, Z., Ruzzo, W.L., and Breaker, R.R. (2005). 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *Rna* 11(5), 774--784. doi: 10.1261/rna.7286705.

Bastián, F., Cohen, A., Piccoli, P., Luna, V., Bottini*, R., Baraldi, R., *et al.* (1998). Production of indole-3-acetic acid and gibberellins A1 and A3 by *Acetobacter diazotrophicus* and *Herbaspirillum seropedicae* in chemically-defined culture media. *Plant Growth Regulation* 24(1), 7-11. doi: 10.1023/a:1005964031159.

Batista, M.B., Teixeira, C.S., Sfeir, M.Z.T., Alves, L.P.S., Valdameri, G., Pedrosa, F.O., *et al.* (2018). PHB Biosynthesis Counteracts Redox Stress in *Herbaspirillum seropedicae*. *Front Microbiol* 9, 472. doi: 10.3389/fmicb.2018.00472.

Bobrovskyy, M., and Vanderpool, C.K. (2013). Regulation of Bacterial Metabolism by Small RNAs Using Diverse Mechanisms. *Annual Review of Genetics* 47(1), 209--232. doi: 10.1146/annurev-genet-111212-133445.

Bonato, P., Batista, M.B., Camilios-Neto, D., Pankievicz, V.C.S., Tadra-Sfeir, M.Z., Monteiro, R.A., *et al.* (2016). RNA-seq analyses reveal insights into the function of respiratory nitrate reductase of the diazotroph *Herbaspirillum seropedicae*. *Environmental microbiology* 18(8), 2677--2688. doi: 10.1111/1462-2920.13422.

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Current Opinion in Microbiology* 10(2), 102--109. doi: 10.1016/j.mib.2007.03.012.

Burenina, O.Y., Elkina, D.A., Hartmann, R.K., Oretskaya, T.S., and Kubareva, E.A. (2015). Small noncoding 6S RNAs of bacteria. *Biochemistry (Moscow)* 80(11), 1429--1446. doi: 10.1134/S0006297915110048.

Carrier, M. C., Lalaouna, D., and Massé, E. (2018). Broadening the definition of bacterial small RNAs: Characteristics and mechanisms of action. *Annu Rev Microbiol* 72, 141-161. doi: 10.1146/annurev-micro-090817-062607.

Carrier, M. C., Morin, C., and Masse, E. (2018). On the Prowl: An In Vivo Method to Identify RNA Partners of a sRNA. *Methods in enzymology* 612, 251-268. doi: 10.1016/bs.mie.2018.09.006

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., and McQuillan, J.A. (2012). Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28(4)**,** 464-469. doi: 10.1093/bioinformatics/btr703.

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. (2009). DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics* 25(1)**,** 119--120. doi: 10.1093/bioinformatics/btn578.

Chao, Y., Li, L., Girodat, D., Forstner, K.U., Said, N., Corcoran, C., *et al.* (2017). *In Vivo* Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Mol Cell* 65(1)**,** 39-51. doi: 10.1016/j.molcel.2016.11.002.

Chubatsu, L.S., Monteiro, R.A., de Souza, E.M., de Oliveira, M.A.S., Yates, M.G., Wassem, R., *et al.* (2012). Nitrogen fixation control in *Herbaspirillum seropedicae*. *Plant and Soil* 356(1-2)**,** 197--207. doi: 10.1007/s11104-011-0819-6.

Coenye, T., Drevinek, P., Mahenthiralingam, E., Shah, S.A., Gill, R.T., Vandamme, P., *et al.* (2007). Identification of putative noncoding RNA genes in the *Burkholderia cenocepacia* J2315 genome. *FEMS Microbiology Letters* 276(1)**,** 83--92. doi: 10.1111/j.1574-6968.2007.00916.x.

Dambach, M., Sandoval, M., Updegrove, T.B., Anantharaman, V., Aravind, L., Waters, L.S., *et al.* (2015). The Ubiquitous yybP-ykoY Riboswitch Is a Manganese-Responsive Regulatory Element. *Molecular Cell* 57(6)**,** 1099--1109. doi: 10.1016/j.molcel.2015.01.035.

Dennis, P.P., and Omer, A. (2005). Small non-coding RNAs in Archaea. *Current Opinion in Microbiology* 8(6)**,** 685--694. doi: 10.1016/j.mib.2005.10.013.

Derman, Y., Sderholm, H., Lindstrm, M., and Korkeala, H. (2015). Role of *csp* genes in NaCl, pH, and ethanol stress response and motility in Clostridium botulinum ATCC 3502. *Food Microbiology* 46**,** 463--470. doi: 10.1016/j.fm.2014.09.004.

Ducker, G.S., and Rabinowitz, J.D. (2015). ZMP: A master regulator of one-carbon metabolism. *Molecular Cell* 57(2)**,** 203--204. doi: 10.1016/j.molcel.2015.01.012.

Dutta, T., and Srivastava, S. (2018). Small RNA-mediated regulation in bacteria: A growing palette of diverse mechanisms. *Gene* 656(November 2017)**,** 60--72. doi: 10.1016/j.gene.2018.02.068.

Edwards, A.L., Reyes, F.E., Hroux, A., and Batey, R.T. (2010). Structural basis for recognition of S-adenosylhomocysteine by riboswitches. *Rna* 16(11)**,** 2144--2155. doi: 10.1261/rna.2341610.

Estrada, G.A., Baldani, V.L.D., de Oliveira, D.M., Urquiaga, S., and Baldani, J.I. (2013). Selection of phosphate-solubilizing diazotrophic *Herbaspirillum* and *Burkholderia* strains and their effect on rice crop yield and nutrient uptake. *Plant and Soil* 369(1-2)**,** 115--129. doi: 10.1007/s11104-012-1550-7.

Fuentes, D.N., Caldern, P.F., Acua, L.G., Rodas, P.I., Paredes-Sabja, D., Fuentes, J.A., *et al.* (2015). Motility modulation by the small non-coding RNA SroC in *Salmonella Typhimurium*. *FEMS Microbiology Letters* 362(17)**,** 1--8. doi: 10.1093/femsle/fnv135.

Ghosh, S., Dureja, C., Khatri, I., Subramanian, S., Raychaudhuri, S., and Ghosh, S. (2017). Identification of novel small RNAs in *Burkholderia cenocepacia* KC-01 expressed under iron limitation and oxidative stress conditions. *Microbiology*. doi: 10.1099/mic.0.000566.

Gottesman, S., and Storz, G. (2010). Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. *Cold Spring Harb Perspect Biol*. doi: cshperspect.a003798 [pii]10.1101/cshperspect.a003798.

Gough, C., Galera, C., Vasse, J., Webster, G., Cocking, E.C., and Denarie, J. (1997). Specific flavonoids promote intercellular root colonization of *Arabidopsis thaliana* by *Azorhizobium caulinodans* ORS571. *Mol Plant Microbe Interact* 10(5)**,** 560-570. doi: 10.1094/MPMI.1997.10.5.560.

Greenlee, E.B., Stav, S., Atilho, R.M., Brewer, K.I., Harris, K.A., and Malkowski, S.N.a. (2018). Challenges of ligand identification for the second wave of orphan riboswitch candidates. *RNA Biology* 6286(November)**,** 1--14. doi: 10.1080/15476286.2017.1403002.

Gyaneshwar, P., James, E.K., Reddy, P.M., and Ladha, J.K. (2002). *Herbaspirillum* colonization increases growth and nitrogen accumulation in aluminium-tolerant rice varieties. *New Phytologist* 154(1)**,** 131--145. doi: 10.1046/j.1469-8137.2002.00371.x.

Hankins, J.S., Denroche, H., and Mackie, G.A. (2010). Interactions of the RNA-binding protein Hfq with *cspA* mRNA, encoding the major cold shock protein. *J Bacteriol* 192(10)**,** 2482-2490. doi: JB.01619-09 [pii] 10.1128/JB.01619-09.

Hassan, S., and Mathesius, U. (2012). The role of flavonoids in root-rhizosphere signalling: Opportunities and challenges for improving plant-microbe interactions. *Journal of Experimental Botany* 63(9), 3429--3444. doi: 10.1093/jxb/err430.

Henkin, T.M. (2008). Riboswitch RNAs: Using RNA to sense cellular metabolism. *Genes and Development* 22(24), 3383--3390. doi: 10.1101/gad.1747308.

Herbig, A., and Nieselt, K. (2011). NocoRNAc: Characterization of non-coding RNAs in prokaryotes. *BMC Bioinformatics* 12. doi: 10.1186/1471-2105-12-40.

Hershberg, R., Altuvia, S., and Margalit, H. (2003). A survey of small RNA-encoding genes in *Escherichia coli. Nucleic Acids Research* 31(7), 1813--1820. doi: 10.1093/nar/gkg297.

Jiang, W., Fang, L., and Inouye, M. (1996). The role of the 5'-end untranslated region of the mRNA for CspA, the major cold-shock protein of *Escherichia coli*, in cold-shock adaptation. *Journal of Bacteriology* 178(16), 4919--4925. doi: 10.1128/jb.178.16.4919-4925.1996.

Kahrstrom, C.T. (2012). Bacterial Transcription: Introducing the excludon. *Nat Rev Microbiol* 10(7), 443. doi: 10.1038/nrmicro2822.

Kandel, S., Joubert, P., and Doty, S. (2017). Bacterial Endophyte Colonization and Distribution within Plants. *Microorganisms* 5(4), 77. doi: 10.3390/microorganisms5040077.

Khoo, J., Chai, S, Mohamed, R., Nathan, S., and Firdaus-raih, M. (2012). Computational discovery and RT-PCR validation of novel *Burkholderia* conserved and *Burkholderia pseudomallei* unique sRNAs. *BMC Genomics* 13(Suppl 7), S13. doi: 10.1186/1471-2164-13-S7-S13.

Kim, P.B., Nelson, J.W., and Breaker, R.R. (2015). An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol Cell* 57(2), 317-328. doi: 10.1016/j.molcel.2015.01.001.

Kimelman, A., Levy, A., Sberro, H., Kidron, S., Leavitt, A., Amitai, G., *et al.* (2012). A vast collection of microbial genes that are toxic to bacteria. *Genome Research* 22(4), 802--809. doi: 10.1101/gr.133850.111.

Klassen, G., Pedrosa, F.O., Souza, E.M., Funayama, S., and Rigo, L.U. (1997). Effect of nitrogen compounds on nitrogenase activity in *Herbaspirillum seropedicae* SMR1. *Canadian journal of microbiology* 43(9), 887--891. doi: 10.1139/m97-129.

Klein, R.J., Misulovin, Z., and Eddy, S.R. (2002). Noncoding RNA genes identified in AT-rich hyperthermophiles. *Pnas* 99(11)**,** 7542--7547. doi: 10.1073/pnas.112063799.

Lalaouna, D., Simoneau-Roy, M., Lafontaine, D., and Massé, E. (2013). Regulatory RNAs and target mRNA decay in prokaryotes. *Biochimica Et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1829(6), 742-747. doi: 10.1016/j.bbagrm.2013.02.013.

Li, C., Liu, S., Liu, P., Wang, Y., Xu, C., Tao, J., *et al.* (2018). The YebN Leader RNA Acts as a Mn2+ Sensor Required for the Interaction of *Xanthomonas oryzae* and Rice. *Molecular plant-microbe interactions : MPMI***,** 1--35. doi: 10.1094/MPMI-02-18-0043-R.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al.* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16)**,** 2078-2079. doi: 10.1093/bioinformatics/btp352.

Liu, X., Tian, J., Liu, L., Zhu, T., Yu, X., Chu, X., *et al.* (2017). Identification of an operon involved in fluoride resistance in *Enterobacter cloacae* FRM. *Scientific Reports* 7(1)**,** 1--9. doi: 10.1038/s41598-017-06988-1.

Magalhaes Cruz, L., de Souza, E.M., Weber, O.B., Baldani, J.I., Dobereiner, J., and Pedrosa F. O. (2001). 16S ribosomal DNA characterization of nitrogen-fixing bacteria isolated from banana (Musa spp.) and pineapple (Ananas comosus (L.) Merril). *Appl Environ Microbiol* 67(5)**,** 2375-2379. doi: 10.1128/AEM.67.5.2375-2379.2001.

Marin, A.M., Souza, E.M., Pedrosa, F.O., Souza, L.M., Sassaki, G.L., Baura, V.A., *et al.* (2013). Naringenin degradation by the endophytic diazotroph *Herbaspirillum seropedicae* SmR1. *Microbiology* 159(Pt 1)**,** 167-175. doi: 10.1099/mic.0.061135-0.

Matos, R.G., Casinhas, J., and Barria, C. (2017). The role of ribonucleases and sRNAs in the virulence of foodborne pathogens. *Frontiers in Microbiology* 8(MAY)**,** 1--7. doi: 10.3389/fmicb.2017.00910.

Mehdizadeh Aghdam, E., Sinn, M., Tarhriz, V., Barzegar, A., Hartig, J.S., and Hejazi, M.S. (2017). TPP riboswitch characterization in *Alishewanella tabrizica* and *Alishewanella aestuarii* and comparison with other TPP riboswitches. *Microbiological Research* 195**,** 71-80. doi: https://doi.org/10.1016/j.micres.2016.11.003.

Mercier, E., Holtkamp, W., Rodnina, M.V., and Wintermeyer, W. (2017). Signal recognition particle binds to translating ribosomes before emergence of a signal anchor sequence. *Nucleic acids research* 45(20)**,** 11858--11866. doi: 10.1093/nar/gkx888.

Michaux, C., Verneuil, N., Hartke, A., and Giard, J.C. (2014). Physiological roles of small RNA molecules. *Microbiology (United Kingdom)* 160(PART 6)**,** 1007--1019. doi: 10.1099/mic.0.076208-0.

Narberhaus, F., and Vogel, J. (2009). Regulatory RNAs in prokaryotes: here, there and everywhere. *Mol Microbiol* 74(2)**,** 261-269. doi: 10.1111/j.1365-2958.2009.06869.x.

Nelson, J.W., Atilho, R.M., Sherlock, M.E., Stockbridge, R.B., and Breaker, R.R. (2017). Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. *Mol Cell* 65(2)**,** 220-230. doi: 10.1016/j.molcel.2016.11.019.

Nitzan, M., Rehani, R., and Margalit, H. (2017). Integration of bacterial small RNAs in regulatory networks. *Annual review of biophysics* 46, 131-148. doi: 10.1146/annurev-biophys-070816-034058.

Olivares, F.L., Baldani, V.L.D., Reis, V.M., Baldani, J.I., and Döbereiner, J. (1996). Occurrence of the endophytic diazotrophs *Herbaspirillum* spp. in roots, stems, and leaves, predominantly of Gramineae. *Biology and Fertility of Soils* 21(3)**,** 197-200. doi: 10.1007/bf00335935.

Pankievicz V.C., Camilios-Neto D., Bonato P., Balsanelli E., Tadra-Sfeir M.Z., Faoro H., Chubatsu L.S., Donatti L., Wajnberg G., Passetti F., Monteiro R.A., Pedrosa F.O., and Souza E.M. (2016). RNA-seq transcriptional profiling of *Herbaspirillum seropedicae* colonizing wheat (*Triticum aestivum*) roots. *Plant Mol Biol* 90(6), 589-603. doi: 10.1007/s11103-016-0430-6.

Pedrolli, D., Langer, S., Hobl, B., Schwarz, J., Hashimoto, M., and Mack, M. (2015). The ribB FMN riboswitch from *Escherichia coli* operates at the transcriptional and translational level and regulates riboflavin biosynthesis. *FEBS Journal* 282(16)**,** 3230--3242. doi: 10.1111/febs.13226.

Pedrosa, F.O., Monteiro, R.A., Wassem, R., Cruz, L.M., Ayub, R.A., Colauto, N.B., *et al.* (2011). Genome of *Herbaspirillum seropedicae* strain SmR1, a specialized diazotrophic endophyte of tropical grasses. *PLoS Genet* 7(5)**,** e1002064. doi: 10.1371/journal.pgen.1002064.

Pimentel, J.P., Olivares, F., Pitard, R.M., Urquiaga, S., Akiba, F., and Döbereiner, J. (1991). Dinitrogen fixation and infection of grass leaves by *Pseudomonas rubrisubalbicans* and *Herbaspirillum seropedicae*. *Plant and Soil* 137(1)**,** 61-65. doi: 10.1007/bf02187433.

Polaski, J.T., Webster, S.M., Johnson, J.E., and Batey, R.T. (2017). Cobalamin riboswitches exhibit a broad range of ability to discriminate between methylcobalamin and adenosylcobalamin. *Journal of Biological Chemistry* 292(28)**,** 11650--11658. doi: 10.1074/jbc.M117.787176.

Price, I.R., Gaballa, A., Ding, F., Helmann, J.D., and Ke, A. (2015). Mn(2+)-sensing mechanisms of yybP-ykoY orphan riboswitches. *Mol Cell* 57(6)**,** 1110-1123. doi: 10.1016/j.molcel.2015.02.016.

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. (2003). Comparative Genomics of the Vitamin B12Metabolism and Regulation in Prokaryotes. *Journal of Biological Chemistry* 278(42)**,** 41148--41159. doi: 10.1074/jbc.M305837200.

Roncato-Maccari, L.D., Ramos, H.J., Pedrosa, F.O., Alquini, Y., Chubatsu, L.S., Yates, M.G., *et al.* (2003). Endophytic *Herbaspirillum seropedicae* expresses nif genes in gramineous plants. *FEMS Microbiol Ecol* 45(1)**,** 39-47. doi: 10.1016/S0168-6496(03)00108-9.

Rumble, S.M., Lacroute, P., Dalca, A.V., Fiume, M., Sidow, A., and Brudno, M. (2009). SHRiMP: accurate mapping of short color-space reads. *PLoS Comput Biol* 5(5)**,** e1000386. doi: 10.1371/journal.pcbi.1000386.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., *et al.* (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16(10)**,** 944-945.

Saberi, F., Kamali, M., Najafi, A., Yazdanparast, A., and Moghaddam, M.M. (2016). Natural antisense RNAs as mRNA regulatory elements in bacteria: a review on function and applications. *Cellular \& Molecular Biology Letters* 21(1)**,** 6. doi: 10.1186/s11658-016-0007-z.

Sass, A., Kiekens, S., and Coenye, T. (2017). Identification of small RNAs abundant in *Burkholderia cenocepacia* biofilms reveal putative regulators with a potential role in carbon and iron metabolism. *Scientific Reports* 7(1)**,** 1--13. doi: 10.1038/s41598-017-15818-3.

Sass, A.M.a. (2015). Genome-wide transcription start site profiling in biofilm-grown *Burkholderia cenocepacia* J2315. *BMC Genomics* 16(1)**,** 1--15. doi: 10.1186/s12864-015-1993-3.

Schmidt, M.A., Souza, E.M., Baura, V., Wassem, R., Yates, M.G., Pedrosa, F.O., *et al.* (2011). Evidence for the endophytic colonization of *Phaseolus vulgaris*(common

bean) roots by the diazotroph Herbaspirillum seropedicae. *Brazilian Journal of Medical and Biological Research* 44**,** 182-185.

Sharma, C.M., and Heidrich, N. (2012). Small RNAs and virulence in bacterial pathogens. *RNA Biology* 9(4)**,** 361--363. doi: 10.4161/rna.20517.

Steuten, B., Hoch, P.G., Damm, K., Schneider, S., Khler, K., Wagner, R., *et al.* (2014). Regulation of transcription by 6S RNAs insights from the *Escherichia coli* and *Bacillus subtilis* model systems. *RNA Biology* 11(5). doi: 10.4161/rna.28827.

Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell* 43(6)**,** 880-891. doi: 10.1016/j.molcel.2011.08.022.

Sun, E.I., Leyn, S.A., Kazanov, M.D., Saier, M.H., Jr., Novichkov, P.S., and Rodionov, D.A. (2013). Comparative genomics of metabolic capacities of regulons controlled by cis-regulatory RNA motifs in bacteria. *BMC Genomics* 14**,** 597. doi: 10.1186/1471-2164-14-597.

Sundermeier, T.R., Dulebohn, D.P., Cho, H.J., and Karzai, A.W. (2005). A previously uncharacterized role for small protein B (*SmpB*) in transfer messenger RNA-mediated trans-translation. *Proceedings of the National Academy of Sciences of the United States of America* 102(7)**,** 2316--2321. doi: 10.1073/pnas.0409694102.

Tadra-Sfeir, M.Z., Faoro, H., Camilios-Neto, D., Brusamarello-Santos, L., Balsanelli, E., Weiss, V., *et al.* (2015). Genome wide transcriptional profiling of *Herbaspirillum seropedicae* SmR1 grown in the presence of naringenin. *Frontiers in Microbiology* 6(MAY)**,** 1--8. doi: 10.3389/fmicb.2015.00491.

Tadra-Sfeir, M.Z., Souza, E.M., Faoro, H., Mller-Santos, M., Baura, V.A., Tuleski, T.R., *et al.* (2011). Naringenin regulates expression of genes involved in cell wall synthesis in *Herbaspirillum seropedicae. Applied and Environmental Microbiology* 77(6)**,** 2180--2183. doi: 10.1128/AEM.02071-10.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., *et al.* (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7(3)**,** 562--578. doi: 10.1038/nprot.2012.016.

Trotochaud, A.E., and Wassarman, K.M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. *Nature Structural and Molecular Biology* 12(4)**,** 313--319. doi: 10.1038/nsmb917.

Valverde, C., Livny, J., Schluter, J.P., Reinkensmeier, J., Becker, A., and Parisi, G. (2008). Prediction of *Sinorhizobium meliloti* sRNA genes and experimental detection in strain 2011. *BMC Genomics* 9**,** 416. doi: 10.1186/1471-2164-9-416.

Wang, J.X., Lee, E.R., Morales, D.R., Lim, J., and Breaker, R.R. (2008). Riboswitches that Sense S-adenosylhomocysteine and Activate Genes Involved in Coenzyme Recycling. *Molecular Cell* 29(6)**,** 691--702. doi: 10.1016/j.molcel.2008.01.012.

Wassarman, K.M., and Storz, G. (2000). 6S RNA regulates *E. coli* RNA polymerase activity. *Cell* 101(6)**,** 613-623.

Waters, L.S., and Storz, G. (2009). Regulatory RNAs in bacteria. *Cell* 136(4)**,** 615-628. doi: S0092-8674(09)00125-1 [pii]10.1016/j.cell.2009.01.043.

Winkler, W.C., and Breaker, R.R. (2005). Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* 59**,** 487-517. doi: 10.1146/annurev.micro.59.030804.121336.

Zaratiegui, M., Irvine, D.V., and Martienssen, R.A. (2007). Noncoding RNAs and Gene Silencing. *Cell* 128(4)**,** 763--776. doi: 10.1016/j.cell.2007.02.016.

Zhang, Y., Burkhardt, D.H., Rouskin, S., Li, G.W., Weissman, J.S., and Gross, C.A. (2018). A Stress Response that Monitors and Regulates mRNA Structure Is Central to Cold Shock Adaptation. *Molecular Cell* 70(2)**,** 274--286.e277. doi: 10.1016/j.molcel.2018.02.035.

Zhao, T., Zhang, R., and Wang, M. (2010). Prediction of candidate small non-coding RNAs in by *Agrobacterium* by computational analysis. *Journal of biomedical research* 24(1), 33-42. doi: 10.1016/S1674-8301(10)60006-1.

Zorgani, M.A., Quentin, R., and Lartigue, M.F. (2016). Regulatory RNAs in the less studied streptococcal species: From nomenclature to identification. *Frontiers in Microbiology* 7(JUL). doi: 10.3389/fmicb.2016.01161.

4.8. ADDITIONAL DATA 1. *Herbaspirillum seropedicae* SmR1 PREDICTED sRNAs

sRNA	RFAM	Prediction tools	Category	GC %	Start	End	Length	Strand	Coverage			RPKM		
									CRT	NAR	NIT	CRT	NAR	NIT
Hsnc001	RF00169	nocoRNAc	trans-encoded	65.66	4108907	4109005	98	-	3433.96	2792.20	2297.02	1769.26	3643.21	2067.06
Hsnc002	RF01734	CuffLinks	riboswitch	60.36	2800595	2800705	110	+	39.27	19.64	35.35	20.23	25.62	31.81
Hsnc003		CuffLinks	trans-encoded	62.79	3181155	3181240	85	-	48.71	42.78	34.31	25.09	55.81	30.87
Hsnc004		CuffLinks	cis-encoded	62.18	3160202	3160357	155	+	29.03	1.39	6.97	14.96	1.82	6.27
Hsnc005		CuffLinks	trans-encoded	61.79	3382844	3382966	122	-	288.89	110.95	310.72	148.84	144.77	279.61
Hsnc006	RF01070	CuffLinks	riboswitch	56.38	3382646	3382739	93	-	248.90	184.26	3961.94	85.56	354.06	3785.10
Hsnc007		CuffLinks	trans-encoded	47.44	359683	359916	233	+	88.84	76.79	160.22	45.77	100.19	144.18
Hsnc008		CuffLinks	cis-encoded	54.43	3204507	3204585	78	+	57.23	13.38	0.00	29.49	17.46	0.00
Hsnc009		CuffLinks	trans-encoded	39.62	3303508	3303560	52	+	115.62	155.08	213.92	59.57	202.34	192.51
Hsnc010		CuffLinks	trans-encoded	43.33	3553031	3553120	89	-	38.43	27.91	14.16	19.80	36.42	12.74
Hsnc011		CuffLinks	trans-encoded	57.95	3614107	3614194	87	-	110.48	37.24	10.76	56.92	48.59	9.68
Hsnc012		CuffLinks	cis-encoded	67.72	3630870	3630996	126	-	267.43	33.43	3.71	137.79	43.62	3.34
Hsnc013		CuffLinks	trans-encoded	60.71	3512060	3512227	167	+	198.97	167.50	254.80	102.51	218.55	229.29
Hsnc014		CuffLinks	trans-encoded	49.36	412442	412676	234	+	56.77	46.77	52.62	29.25	61.02	47.35
Hsnc015		CuffLinks	trans-encoded	59.23	4402051	4402283	232	-	31.81	12.10	12.72	16.39	15.79	11.45
Hsnc016		CuffLinks	trans-encoded	47.32	4499669	4499780	111	-	34.38	108.32	326.92	17.71	141.34	294.19

sRNA	RFAM	Prediction tools	Category	GC %	Start	End	Length	Strand	Coverage			RPKM		
									CRT	NAR	NIT	CRT	NAR	NIT
Hsnc017		CuffLinks	trans-encoded	49.57	4587035	4587149	114	-	33.47	23.37	9.47	17.25	30.49	8.53
Hsnc018		CuffLinks	trans-encoded	60.17	4834369	4834486	117	-	185.23	25.54	64.92	95.44	33.32	58.42
Hsnc019		CuffLinks	trans-encoded	48.76	5000211	5000493	282	-	63.06	48.51	23.62	32.49	63.30	21.25
Hsnc020		CuffLinks	trans-encoded	55.22	4909594	4909823	229	-	76.72	39.62	30.50	39.53	51.69	27.44
Hsnc021		CuffLinks	trans-encoded	58.10	4925909	4926013	104	+	34.27	47.08	34.27	17.66	61.42	30.84
Hsnc022		CuffLinks	trans-encoded	74.55	4929103	4929157	54	-	202.67	96.00	324.00	104.42	125.26	291.56
Hsnc023		CuffLinks	cis-encoded	45.90	664741	664801	60	+	58.20	37.80	97.80	29.99	49.32	88.01
Hsnc024		CuffLinks	cis-encoded	62.16	539844	539917	73	-	23.67	11.34	5.42	12.20	14.80	4.88
Hsnc025		CuffLinks	trans-encoded	48.12	741007	741139	132	+	609.55	161.18	146.73	314.05	210.31	132.04
Hsnc026		CuffLinks	trans-encoded	50.00	792569	792670	101	+	108.36	40.28	6.42	55.83	52.55	5.77
Hsnc027		CuffLinks	trans-encoded	51.58	1155828	1155922	94	+	62.43	10.34	60.89	32.16	13.49	54.80
Hsnc028		CuffLinks	trans-encoded	47.62	1336439	1336564	125	+	23.62	64.51	218.88	12.17	84.17	196.97
Hsnc029	RF01750	CuffLinks	riboswitch	61.00	1381288	1381387	99	+	67.27	8.00	50.55	34.66	10.44	45.49
Hsnc030		CuffLinks	trans-encoded	52.33	205713	205798	85	+	132.99	31.34	1.27	68.52	40.89	1.14
Hsnc031		CuffLinks	cis-encoded	49.43	3781409	3781584	175	+	31.06	45.67	332.02	16.00	59.59	298.78
Hsnc032		CuffLinks	cis-encoded	60.19	1867577	1867679	102	+	44.47	6.00	56.82	22.91	7.83	51.13
Hsnc033		CuffLinks	cis-encoded	60.71	1880603	1880826	223	+	1.13	2.42	5.49	0.58	3.16	4.94

sRNA	RFAM	Prediction tools	Category	GC %	Start	End	Length	Strand	Coverage			RPKM		
									CRT	NAR	NIT	CRT	NAR	NIT
Hsnc034		CuffLinks	trans-encoded	62.07	2011544	2011601	57	+	79.58	10.11	232.42	41.00	13.19	209.15
Hsnc035	RF00050	CuffLinks	riboswitch	65.09	2225947	2226115	168	+	83.57	261.00	304.29	43.06	340.55	273.82
Hsnc036		CuffLinks	trans-encoded	64.71	2287397	2287464	67	+	149.37	32.78	266.51	76.96	42.77	239.83
Hsnc037		CuffLinks	trans-encoded	60.49	2519484	2519726	242	+	273.42	122.13	76.02	140.87	159.36	68.41
Hsnc038		CuffLinks	trans-encoded	63.36	2526174	2526304	130	+	41.54	9.14	9.97	21.40	11.92	8.97
Hsnc039		CuffLinks	trans-encoded	54.60	2578050	2578397	347	+	528.80	194.52	21.99	272.45	253.81	19.79
Hsnc040		nocoRNAc	trans-encoded	46.02	60476	60588	112	-	10.93	41.46	26.68	5.63	54.10	24.01
Hsnc041		nocoRNAc	trans-encoded	52.22	328428	328517	89	+	3.64	2.83	9.71	1.88	3.69	8.74
Hsnc042		nocoRNAc/Cufflinks	trans-encoded	49.21	417221	417409	188	+	15.32	90.57	320.17	7.89	118.18	288.12
Hsnc043		nocoRNAc/Cufflinks	trans-encoded	40.00	457262	457311	49	+	11.02	5.14	13.22	5.68	6.71	11.90
Hsnc044		nocoRNAc/Cufflinks	cis-encoded	57.92	551015	551235	220	+	43.04	46.64	17.18	22.17	60.85	15.46
Hsnc045		nocoRNAc/Cufflinks	trans-encoded	55.03	809023	809191	168	+	53.57	58.07	19.71	27.60	75.77	17.74
Hsnc046		nocoRNAc	trans-encoded	35.00	831492	831611	119	-	3.03	24.81	61.41	1.56	32.37	55.26
Hsnc047		nocoRNAc	trans-encoded	59.46	935356	935540	184	+	17.02	4.89	50.48	8.77	6.38	45.42
Hsnc048		nocoRNAc	trans-encoded	58.97	957549	957626	77	+	23.84	12.16	11.69	12.29	15.86	10.52
Hsnc049		nocoRNAc	trans-encoded	65.22	1067480	1067548	68	-	4.76	1.06	10.59	2.45	1.38	9.53
Hsnc050	RF00013	nocoRNAc/Cufflinks	trans-encoded	53.67	1383620	1383796	176	-	15496.77	2919.27	266.11	7984.33	3809.00	239.47

sRNA	RFAM	Prediction tools	Category	GC %	Start	End	Length	Strand	Coverage			RPKM		
									CRT	NAR	NIT	CRT	NAR	NIT
Hsnc051		nocoRNAc	trans-encoded	53.19	1433153	1433246	93	+	25.16	7.74	20.13	12.96	10.10	18.11
Hsnc052		nocoRNAc	trans-encoded	55.5	1596981	1597171	190	-	25.39	28.23	606.88	13.08	36.84	546.13
Hsnc053		nocoRNAc	trans-encoded	51.35	1624918	1624991	73	+	35.01	17.26	26.14	18.04	22.52	23.52
Hsnc054		nocoRNAc	cis-encoded	44.51	1636668	1636831	163	+	31.36	33.35	326.21	16.16	43.51	293.55
Hsnc055		nocoRNAc	trans-encoded	48.86	1725034	1725121	87	+	14.07	12.00	51.72	7.25	15.66	46.55
Hsnc056		nocoRNAc/Cufflinks	trans-encoded	51.55	2032380	2032540	160	+	51.08	39.15	30.38	26.32	51.08	27.33
Hsnc057		nocoRNAc	cis-encoded	64.15	2124304	2124356	52	+	2.77	7.62	20.08	1.43	9.94	18.07
Hsnc058		nocoRNAc	trans-encoded	67.42	2140728	2141037	309	-	156.12	59.30	98.21	80.44	77.37	88.38
Hsnc059	RF02278	nocoRNAc	trans-encoded	49.46	2177255	2177347	92	+	11.74	15.65	130.70	6.05	20.42	117.61
Hsnc060		nocoRNAc	trans-encoded	57.14	2474001	2474042	41	+	7.90	7.90	22.83	4.07	10.31	20.54
Hsnc061		nocoRNAc	trans-encoded	59.31	2815035	2815179	144	-	119.50	111.75	307.75	61.57	145.81	276.94
Hsnc062		nocoRNAc	trans-encoded	59.06	2939684	2939832	148	+	23.59	19.95	19.22	12.16	26.03	17.29
Hsnc063	RF02278	nocoRNAc	trans-encoded	43.48	2972997	2973111	114	+	2.84	38.21	74.21	1.46	49.86	66.78
Hsnc064		nocoRNAc/Cufflinks	trans-encoded	52.53	3173522	3173679	157	-	30.96	81.63	1.61	15.95	106.51	1.44
Hsnc065		nocoRNAc	trans-encoded	61.88	3317231	3317432	201	+	3.94	4.48	11.28	2.03	5.84	10.15
Hsnc066		nocoRNAc	trans-encoded	59.43	3325609	3325783	174	-	2.48	3.31	6.00	1.28	4.32	5.40
Hsnc067		nocoRNAc/Cufflinks	cis-encoded	55	3633077	3633376	299	-	3263.00	1405.81	1114.92	1681.18	1834.27	1003.30

sRNA	RFAM	Prediction tools	Category	GC %	Start	End	Length	Strand	Coverage			RPKM		
									CRT	NAR	NIT	CRT	NAR	NIT
Hsnc068		nocoRNAc	cis-encoded	52.23	3750872	3751432	560	+	94.18	62.16	424.22	48.52	81.11	381.75
Hsnc069		nocoRNAc	cis-encoded	53.32	3776244	3776620	376	+	3.26	1.91	10.63	1.68	2.50	9.56
Hsnc070		nocoRNAc/Cufflinks	trans-encoded	61.39	3779834	3779991	157	-	58.24	31.64	173.81	30.01	41.29	156.41
Hsnc071		nocoRNAc/Cufflinks	trans-encoded	73.17	3791718	3791799	81	-	24.00	2.22	1.78	12.37	2.90	1.60
Hsnc072		nocoRNAc	trans-encoded	55.45	3879803	3880105	302	+	0.72	1.31	32.54	0.37	1.71	29.29
Hsnc073		nocoRNAc	trans-encoded	55.49	3928375	3928556	181	+	304.91	320.62	339.91	157.10	418.34	305.88
Hsnc074		nocoRNAc	trans-encoded	47.33	3989988	3990380	392	+	5.79	6.43	9.73	2.98	8.39	8.76
Hsnc075		nocoRNAc	cis-encoded	45.65	3995857	3996132	275	+	2.62	6.02	16.76	1.35	7.86	15.08
Hsnc076		nocoRNAc	trans-encoded	47.19	3998440	3998528	88	-	9.00	5.32	32.73	4.64	6.94	29.45
Hsnc077		nocoRNAc	trans-encoded	47.22	4004964	4005107	143	-	5.54	9.06	21.15	2.85	11.83	19.03
Hsnc078		nocoRNAc	cis-encoded	48.75	4010423	4010502	79	+	351.80	1145.16	928.71	181.25	1494.19	835.73
Hsnc079		nocoRNAc	trans-encoded	49.58	4074932	4075171	239	-	9.64	15.67	194.31	4.97	20.44	174.86
Hsnc080		nocoRNAc	cis-encoded	61.48	4105880	4106014	134	-	9.94	17.73	0.54	5.12	23.14	0.48
Hsnc081		nocoRNAc	trans-encoded	50.62	4127472	4127552	80	-	24.75	214.20	121.05	12.75	279.48	108.93
Hsnc082		nocoRNAc/Cufflinks	trans-encoded	45.86	4136551	4136707	156	-	545.31	266.54	174.46	280.96	347.77	157.00
Hsnc083	RF00023	nocoRNAc/Cufflinks	trans-encoded	52.34	4239191	4239574	383	+	57089.04	17093.33	29174.85	29413.71	22303.00	26254.09
Hsnc084		nocoRNAc	trans-encoded	34.57	4319058	4319138	80	-	4.05	4.50	12.60	2.09	5.87	11.34

sRNA	RFAM	Prediction tools	Category	GC %	Start	End	Length	Strand	Coverage			RPKM		
									CRT	NAR	NIT	CRT	NAR	NIT
Hsnc085		nocoRNAc	cis-encoded	40.00	4388096	4388200	104	-	2.42	2.08	54.35	1.25	2.71	48.91
Hsnc086	RF02278	nocoRNAc	trans-encoded	47.42	4408980	4409076	96	-	4.13	15.00	52.50	2.13	19.57	47.24
Hsnc087		nocoRNAc	trans-encoded	72.12	4417512	4417615	103	-	25.51	5.59	14.68	13.15	7.30	13.21
Hsnc088		nocoRNAc	trans-encoded	61.22	4428208	4428599	391	-	13.81	34.25	14.92	7.12	44.69	13.42
Hsnc089		nocoRNAc	cis-encoded	65.88	4473739	4474078	339	-	9.66	5.31	8.81	4.98	6.93	7.93
Hsnc090		nocoRNAc	trans-encoded	53.90	4474802	4474942	140	-	16.71	9.77	71.74	8.61	12.75	64.56
Hsnc091		nocoRNAc	cis-encoded	58.33	4564991	4565134	143	-	5.54	2.77	9.06	2.85	3.61	8.16
Hsnc092		nocoRNAc	trans-encoded	49.46	4565726	4565818	92	+	5.87	1.96	34.83	3.02	2.55	31.34
Hsnc093		nocoRNAc	cis-encoded	49.36	4896557	4896712	155	+	107.07	104.52	401.81	55.17	136.37	361.58
Hsnc094		nocoRNAc	cis-encoded	50.32	5019552	5019706	154	+	246.86	249.43	753.19	127.19	325.45	677.79
Hsnc095		nocoRNAc	trans-encoded	53.78	5026868	5026986	118	-	2.44	8.85	245.59	1.26	11.54	221.01
Hsnc096		nocoRNAc/Cufflinks	trans-encoded	58.20	5239869	5240191	322	+	10.62	13.86	52.66	5.47	18.09	47.39
Hsnc097		nocoRNAc	trans-encoded	58.80	5335529	5335812	283	+	33.84	5.60	36.51	17.43	7.30	32.85
Hsnc098		nocoRNAc	cis-encoded	55.56	5433936	5434088	152	+	2.61	2.37	8.53	1.34	3.09	7.67
Hsnc099		nocoRNAc	trans-encoded	56.19	747547	747651	104	-	5.88	3.12	93.12	3.03	4.06	83.79
Hsnc100		nocoRNAc	trans-encoded	56.44	1943525	1943726	201	+	171.94	71.28	35.46	88.59	93.01	31.91
Hsnc101		nocoRNAc	trans-encoded	53.66	3815799	3815921	122	+	18.59	31.28	49.28	9.58	40.81	44.35

sRNA	RFAM	Prediction tools	Category	GC %	Start	End	Length	Strand	Coverage			RPKM		
									CRT	NAR	NIT	CRT	NAR	NIT
Hsnc102		Cufflinks	trans-encoded	56.59	2234033	2234214	181	+	60.07	34.41	62.85	30.95	44.90	56.56
Hsnc103		nocoRNAc	cis-encoded	51.58	5058503	5058597	94	+	1.53	1.91	6.51	0.79	2.50	5.86
Hsnc104		nocoRNAc	trans-encoded	42.00	808782	808981	199	-	20.80	41.79	27.14	10.72	54.53	24.42
Hsnc105		nocoRNAc	trans-encoded	57.32	4872868	4872949	81	+	8.00	0.44	16.00	4.12	0.58	14.40
Hsnc106		nocoRNAc	trans-encoded	60.75	1788675	1788860	185	-	9.73	5.84	27.24	5.01	7.62	24.52
Hsnc107	RF02278	nocoRNAc/Cufflinks	trans-encoded	43.88	4127685	4127782	97	-	170.35	1688.29	489.90	87.77	2202.84	440.85
Hsnc108		nocoRNAc	trans-encoded	51.94	1915816	1915944	128	+	3.38	13.78	115.31	1.74	17.98	103.77
Hsnc109	RF00059	Cufflinks	riboswitch	58.36	462470	462762	292	+	42.04	19.60	118.85	21.66	25.58	106.95
Hsnc110	RF00442	nocoRNAc	riboswitch	64.36	5118406	5118506	100	-	1.08	2.16	1.08	0.56	2.82	0.97
Hsnc111	RF00010	nocoRNAc	trans-encoded	61.19	357069	357403	334	+	2388.50	828.65	612.54	1230.62	1081.20	551.22
Hsnc112	RF00080	nocoRNAc	riboswitch	63.04	576481	576664	183	+	6.49	11.80	4.92	3.34	15.40	4.43
Hsnc113	RF01057	nocoRNAc	riboswitch	65.56	4908117	4908206	89	-	91.01	30.34	212.76	46.89	39.58	191.46
Hsnc114	RF01766	nocoRNAc	cis-encoded	51.47	1599563	1599935	372	-	267.97	318.00	7452.97	138.06	414.92	6706.84
Hsnc115	RF00174	nocoRNAc	riboswitch	63.97	3038673	3038919	246	-	50.63	25.61	38.63	26.09	33.42	34.77
Hsnc116	RF00442	nocoRNAc	riboswitch	60.00	5119175	5119274	99	+	2.06	10.44	0.00	4.00	8.00	0.00
Hsnc117	RF01766	nocoRNAc	cis-encoded	48.97	3448131	3448518	387	+	138.60	140.93	2429.95	71.23	183.41	2181.05

5. ARTIGO 2: NETWORK INTERACTIONS OF sRNAs-mRNAs FROM *Herbaspirillum seropedicae* SmR1

Tatiane Dobrzanski[1], Fabrício Martins Lopes[2], Vania Pobre[3], Alexandre Rossi Paschoal[2], Cecilia Maria Arraiano[3], Maria Berenice Reynaud Steffens[1]

[1]Department of Biochemistry and Molecular Biology, Universidade Federal do Paraná (UFPR), Av. Coronel. Francisco H. dos Santos, 81.531-980. Curitiba, Brazil.

[1]Department of Computacional Science, Universidade Tecnológica Federal do Paraná (UTFPR), Av. Alberto Carazzai, 86.300-000. Cornélio Procópio, Brazil.

[3]Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Av. da República, 2780-157, Oeiras, Portugal

5.1. ABSTRACT

The sRNAs are the major post-transcriptional regulators in bacteria and are involved in several processes such as adaptation to stress, virulence, motility and *quorum sensing*. Trans-encoded (traRNA) sRNAs can interact by imperfect base pairing with dozens of target mRNAs and regulate it in a post-transcriptional manner. *Herbaspirillum seropedicae* SmR1 is a diazotrophic bacterium belonging to the Betaproteobacteria group which express 77 traRNAs among which only four of them were identified as Betaproteobacteria Toxic Small RNA, but without known targets. This work aimed to infer sRNAs-mRNAs interaction networks in the *H. seropedicae* SmR1. We applied two different approaches for network inference. First, we use expression data from total RNA sequencing as input in five methods that use expression data for network inference and second, we used the sRNAs sequences to search for base pairing regions with mRNAs. Finally, we constructed a consensus network with the common sRNA-mRNA interactions in both methodologies. The sRNA-mRNAs involved mRNAs of all GOG's functional categories and we saw that many genes do not present functional annotation and therefore were grouped in S category (no functional prediction). Various mRNAs coding for nitrogen metabolism protein appeared to be regulated

by sRNAs. Using this methodology, we could define a sRNA-mRNA network interaction with more accurate results which may be more suitable for experimental validation.

Keywords: *Herbaspirillum seropedicae* SmR1, sRNA-mRNA network, Aracne, CLR, DimReduction, Genie3, MRNET, IntaRNA, TargetRNA2

5.2. INTRODUCTION

Small non-coding RNAs (sRNAs) are regulatory molecules with a size ranging from 50 ~ 500 bp (Liu and Camilli, 2010; Gottesman and Storz, 2011) that participate in a series of essential regulatory processes such as adaptation to stress, virulence, motility, and premature termination of transcription (Hoe et al, 2013; Valverde *et al.*, 2008; Silva *et al.*, 2018). These molecules generally modulate gene expression by two mechanisms: (i) interacting with proteins, modifying its activity, or (ii) interacting with mRNAs targets by base pairing, resulting in translation induction/inhibition/termination, or in the cleavage/degradation of RNAs by ribonucleases (Saramago *et al.*, 2014, Dressaire *et al.*, 2018). This pairing is often stabilized by the RNA chaperone, Hfq (Massé *et al.*, 2003; Majdalani *et al.*, 2005; Storz *et al.*, 2011).

The sRNAs can be transcribed *in trans* or *in cis* relative to their targets. Trans-encoded sRNAs (traRNA) exhibit imperfect complementarity of base pairs with their targets, since they are expressed from genomic regions remote from their targets (Andrade *et al.*, 2012, Richards and Vanderpool, 2011). Instead, cis-encoded sRNAs (caRNA) exhibit high base pairs complementarity with their targets since they are located in the complementary strand (Richards and Vanderpool, 2011). The caRNAs have the capacity to act only on single target whereas the traRNAs have the ability to interact with dozens of different targets (Zhan *et al.*, 2016). The mechanism of action of sRNAs usually involves blocking the ribosome binding site (RBS) or the translation start codon of the mRNA target. However, there may also be interaction of the sRNA with the coding sequence of the mRNA (Valverde *et al.*, 2008; Gottesman and Storz, 2011; Vogel and Wagner, 2007).

Most sRNAs found in Bacteria and Archaea are trans-encoded which regulate a varied number of targets (Waters and Storz, 2009; Babski *et al.*, 2014) and the large amount of interactions based on imperfect base pairing hamper detection target mRNAs for the traRNAs. Advanced techniques such as *microarray*, cDNA sequencing (also known RNA

deep sequencing or RNA-seq) and the bioinformatics analysis allows the study of the gene expression of the organisms and contribute to the discovery of sRNAs and their targets (Lopes *et al.*, 2009; Modi *et al.*, 2011; Marbach *et al.*, 2012; Carrier *et al.*, 2018; Lalaouna *et al.*, 2013)

The inference of interaction networks is an approach used to predict possible targets and functions for non-coding RNAs and this network can be inferred from gene expression data sets. Several algorithms have been developed for interaction networks inference and were applied in studies focused on bacteria. Modi *et al.* (2011) used a compendium of expression data as input to the Context Likelihood of Relatedness (CLR) algorithm to infer the sRNA regulatory network in *Escherichia coli*. The Weighted Gene Co-expression Network Analysis (WGCNA) algorithm was used by Subramanian *et al.* (2018) to identify functions of sRNAs in *Staphylococcus aureus*.

Another methodology widely used on the detection of sRNA targets is based on the search of regions of pairing between sRNA-mRNA in the genome of the organism. This approach was used in the development of the tools TargetRNA2 (Kery *et al.*, 2014) and IntaRNA (Busch *et al.*, 2008).

Herbaspirillum seropedicae SmR1 is a nitrogen-fixing bacterium belonging to the group of β-Proteobacteria that endophytically colonizes economically important plants such as cereal crops (Baldani *et al.*, 1986; Baldani and Baldani, 2005). It has been shown that 117 sRNAs are expressed in this bacterium of which only 20 have known identity (Dobrzanski *et al.*, 2019 submitted). In this bacterium, 77 traRNAs probably interact with mRNAs, including 4 traRNAs that were identified as Betaproteobacteria Toxic Small RNA.

In this work, we applied different approaches to infer targets and functions for traRNAs of *H. seropedicae* SmR1. We have inferred interaction networks using two approaches: (i) running algorithms that use RNA-seq data as input and (ii) using sRNAs targets searching tools which find sRNA-mRNA base pairing regions on the genome. A consensus network which establishes relationships between traRNAs and mRNAs was obtained and confirms the complex interaction multiplicity of the traRNAs and their mRNAs targets. This bioinformatics approach represents an important step before the experimental validation of targets and functions of traRNAs.

5.3. MATERIALS AND METHODS

5.3.1. RNA -Seq data analysis

Eleven data sets (Additional data 1) of RNA sequencing (RNA-seq) of *H. seropedicae* SmR1 (NC_014323.1) cultured in different conditions were used for sRNA-mRNA network inferences. Small non-coding RNA (sRNA) and mRNAs expression values in PRKM were obtained with Artemis (Rutherford *et al.*, 2000).

5.3.2. sRNA-mRNA network inferences based on expression data

We used two approaches for inference of sRNAs-mRNAs from *H. seropedicae* SmR1 (Figure 1). The first approach is based on inference networks based on expression data and is described in this topic, while the second approach is based on base pairing between sRNA-mRNAs described in next topic.

For inference of networks based on expression data, a data matrix with expression values of 77 *trans* encoded sRNAs and of 4,726 mRNAs from the 11 the RNA-seq experimental conditions were used as input for inference of interaction networks. The following network inference methods were adopted: (i) Algorithm for the Reconstruction of Accurate Cellular Networks (Aracne) (Margolin *et al.*, 2006); (ii) Context Likelihood Relatedness (CLR) (Faith *et al.*, 2007); (iii) Minimum Redundancy / Maximum Relevance Networks (MRNET) (Meyer *et al.*, 2008); (iv) Gene Network Inference with Ensemble of Trees (GENIE3) (Irrthum *et al.*, 2010); and (v) DimReduction - Sequential Forward Floating Selection (SFFS) (Lopes *et al.*, 2008).

The Minet package (Meyer *et al.*, 2008) available in the Bioconductor for software R was used for the Aracne, MRNET and CLR networks inferences, while the Genie3 package (Irrthum *et al.*, 2010) was used for Genie3. We used a threshold of 0.05 for interactions in Genie3. The default parameters were maintained in all inferences and only sRNAs-mRNAs interactions were considered to form networks.

5.3.3. sRNA-mRNA network inferences based on base paring

The tools TargetRNA2 (Kery *et al.*, 2014) e IntaRNA (Busch *et al.*, 2008) were used to search for sRNA targets in the *H. seropedicae* SmR1 genome. The default parameters of these

tools were adopted. The TargetRNA2 tool establishes a P-value threshold less than 0.05 to consider an mRNA as the target of a sRNA. The output from IntaRNA tool lists all probable target mRNAs of each sRNA. In our case we considered as sRNAs targets only the top 100 target mRNAs whose interaction with a sRNA had a p-value less than 0.05. All sRNA-mRNA interactions were put together to compose a single list.

5.3.4. Consensus network inference

All the networks of sRNA-mRNA predicted for *H. seropedicae* SmR1 were used to create a consensus network. This network was composed of: (i) common interactions found in at least 2 networks inferred based on expression data; (ii) the interactions resulted from the TargetRNA and IntaRNA tools. The mRNAs functional categories which appeared in the sRNA-mRNAs interactions were identified using the Cluster of Orthologous Groups - COG (Tatusov et al, 2000). The Cytoscape 3.7.1 software (Shannon et al, 2003) was used for graphical visualization of sRNAs-mRNAs networks.

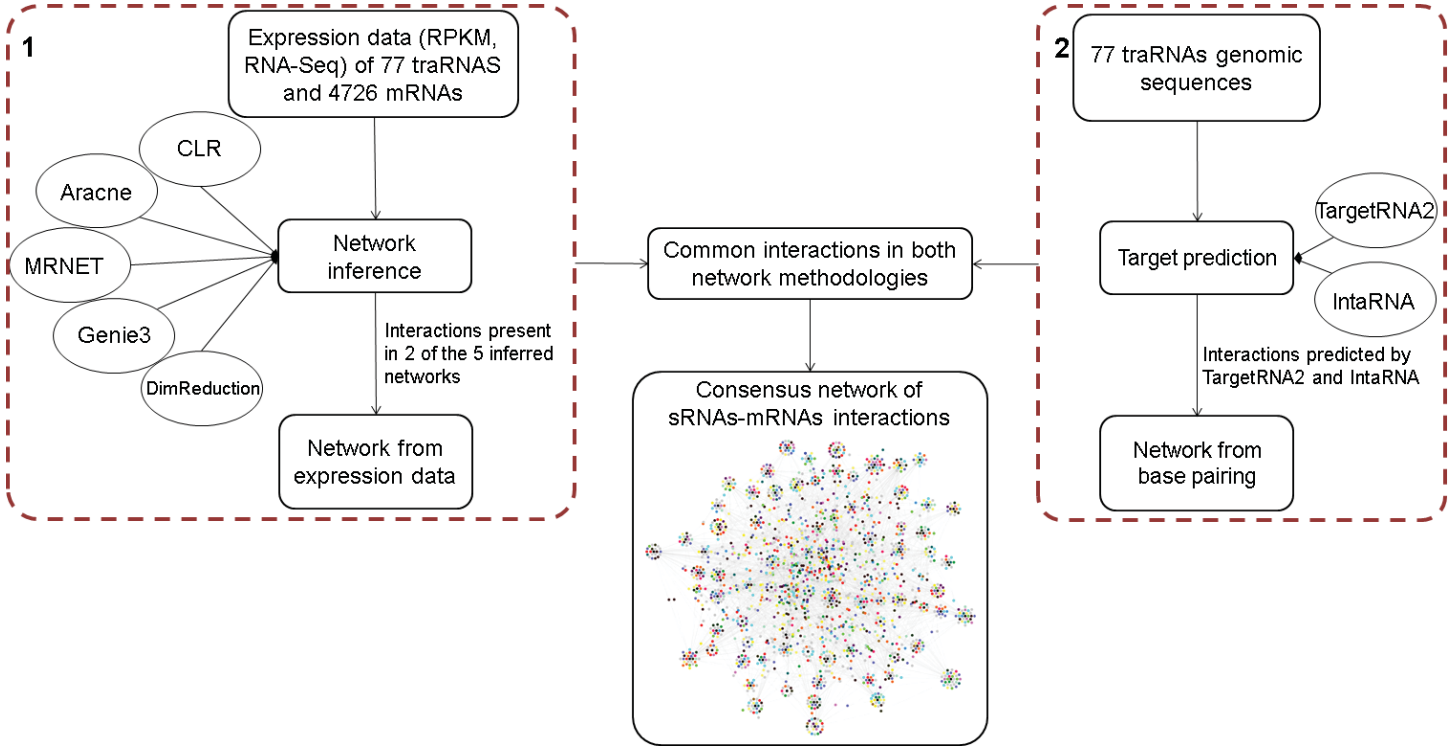


FIGURE 1: Workflow for sRNAs-mRNAs network inference in *H. seropedicae* SmR1. Two network inference approaches were used to identify traRNA-mRNAs interactions on the H. *seropedicae* SmR1. (1) A data matrix with expression values for 77 traRNAs and for 4,726 mRNAs obtained from 11 the RNA-seq experimental conditions was used as input for the interaction networks inference. The following network inference methods were adopted: Aracne, CLR, MRNET, GENIE3 e DimReduction. The common interactions found in at least 2 networks based on expression data were selected to compose the network predicted based on expression data, called HsNetEx. (2) The sequence of the 77 traRNAs were used to search for mRNAs targest through the tools TargetRNA2 and IntaRNA. These tools predict possible binding regions sRNAs on mRNA targets. The total list of interactions predicted by these two tools makes up the base pairing network, called HsNetBP. A consensus network was created by filtering out the common interactions in both HsNetEx and HsNetBP networks.

5.4. RESULTS

5.4.1. sRNA-mRNA Network based on expression data

We have recently identified 81 traRNAs in the *H. seropedicae* SmR1 genome, but only eight of them have a defined identity: 4.5S RNA (Hsnc001), 6S RNA (Hsnc050) and tmRNA (Hsnc083), that interact directly with proteins, RNase P RNA (Hsnc111) which is the catalytic part of the ribozyme for 5'-end pre-tRNA maturation and four Betaproteobacteria Toxic sRNAs (Hsnc059, Hsnc063, Hsnc086, Hsnc0107), also called tsRNAs, that do not yet have known targets. In this work, we constructed sRNA-mRNA interaction networks for the 77 traRNA previously described in *H. seropedicae* SmR1 that possibly interact with mRNAs.

We infered interaction networks using expression data as input to the methods: Aracne, *CLR*, MRNET, GENIE3 and DimReduction. All sRNAs-mRNA interactions predicted by all five methods are available in [1]Additional data 2. The number of sRNA-mRNA interactions (output) was different for each method (Figure 2). The CLR and MRNET networks had the highest number of common sRNA-mRNA interactions (123,142). The Aracne network was the most similar among the others, since only 5 interactions were exclusively predicted by it. The CLR network presented the highest number of unique interactions (14,755), followed by MRNET (4,975), Genie3 (2,998) and DimReduction (698). Five sRNAs-mRNA interactions appeared all networks: Hsnc007-*motB*, Hsnc028-*nuoH*, Hsnc028-HSERO_RS17530, Hsnc070-HSERO_RS16510 and Hsnc095-HSERO_RS09355.

When we applied a filter to select the common interactions in at least 2 networks we got 140,333 sRNA-mRNA interactions which compose a new network based on expression data, called HsNetEx.

[1] Additional data provided by email

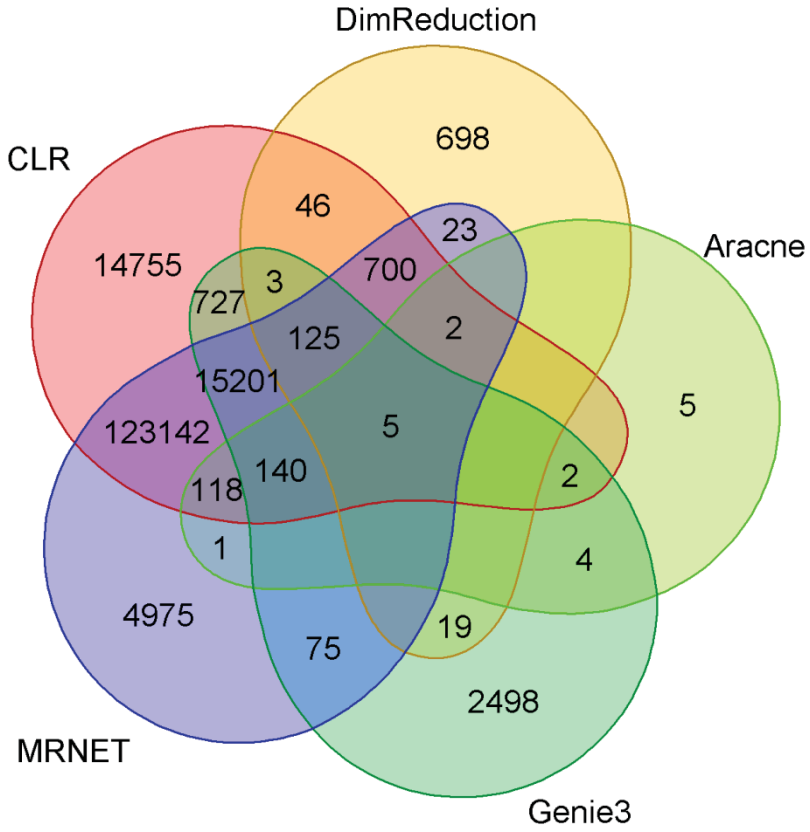


FIGURE 2: Venn diagram showing the intersection of sRNAs-mRNAs interactions inferred using RNA-seq data as input to the Aracne, CLR, DimReduction, MRNET, and Genie3 methods.

5.4.2. The sRNAs targets of *H. seropedicae* SmR1

The matching site for a sRNA in a mRNA sequence is an important criterion to identify an mRNAs target of an sRNAs. The traRNAs usually aligns to 5' non-coding regions closely upstream to the RBS, blocking the mRNA binding sequence of the ribosome (Gottesman and Storz, 2011; Vogel and Wagner, 2007). The targetRNA2 and IntaRNA tools search by default for base pairing regions of sRNAs overlapping to or nearby RBS (Kery *et al.*, 2014; Busch *et al.*, 2008). These two tools consider a minimum number of 7 nucleotides as a hybridization seed corresponding to an initial interaction between the sRNA and mRNA represented by a stretch of consecutive base pairs. When looking for targets from a sRNA, TargetRNA2 focuses its search around the region of 80 nucleotides upstream and 20 nucleotides downstream of the start translation site (start codon) of each mRNA. The IntaRNA tool has a target search region expanded between 150 nucleotides upstream and downstream of the start codon.

We used these tools to predict mRNA targets of the *H. seropedicae* SmR1 sRNAs. An output list with 8,757 sRNA-mRNA interaction was obtained by both tools and compose the network based on base pairing, named HsNetBP (Figure 3). The HsNetBP network consists of 1,403 interactions predicted by the TargetRNA2 and 7,504 interactions predicted by the IntaRNA, and 150 interactions were predicted by both tools.

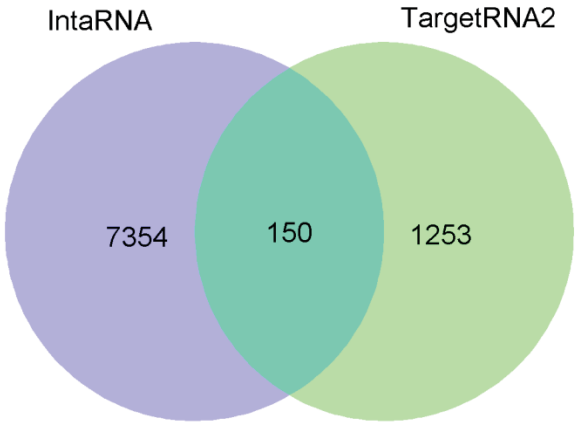


FIGURE 3. Venn diagram showing the common relationships predicted by IntaRNA and TargetRNA2.

5.4.3. sRNA-mRNA consensus network for *Herbaspirillum seropedicae* SmR1

We have selected the common interactions between network based on expression data (HsNetEx) with the interactions present in the base pairing-based network (HsNetBP) to compose a consensus network (Figure 4). This network consists of all 77 traRNAs and 1,988 mRNAs which participate in 3,218 sRNA-mRNA interactions.

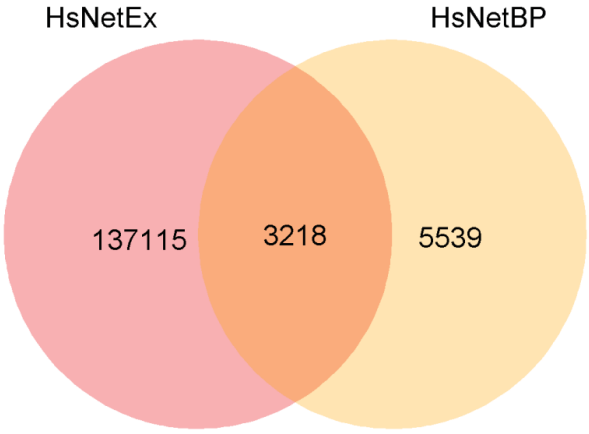


FIGURE 4. Venn diagram showing the common relationships predicted by network-based expression data (HsNetEx) and base pairing-based network (HsNetBP).

The HsNetEx presented 2,881 sRNA-mRNA interactions common to IntaRNA tool and 383 interactions common to TargetRNA2 (Figure 5A). The graphical visualization of the sRNA-mRNA interactions present in the consensus network compared to the interactions predicted by TargetRNA2 and/or IntaRNA is shown in the Figure 5B. Forty-six interactions were identical between HsNetEx and HsNetBP. These interactions have a significant energy hybridization between sRNAs-mRNAs (Additional data 3). The interaction between the Hsnc020 sRNA and the *htrB* mRNA (HSERO_RS21445) showed the most negative hybridization energy according to IntaRNA (-236,397 kcal/mol) and TargetRNA2 (-23.58 kcal/mol). The *htrB* gene codes for lipid A biosynthesis acyltransferase. Expression of the *htrB* gene is essential for responsiveness of *Salmonella typhimurium* and *Campylobacter jejuni* to harsh environments (Phongsisay *et al.*, 2007; Jones *et al.*, 1997). This protein is also essential for viability of *E. coli* above 33 ºC in rich media (Jones *et al.*, 1997). A region with 20 nucleotides Hsnc020 sRNA hybridizes with the mRNA covering a sequence starting at position -1, upstream to AUG, and finishing at position +20 (Additional data 4A).

Among the 46 common interactions between HsNetEx and the predictions of TargetRNA2 and IntaRNA, we observed that the Hsnc107 identified as Betaproteobacteria Toxic sRNA (tsRNA) interacts with the *fliJ* (HSERO_RS10290) and *def* (HSERO_RS10835) mRNAs. The *fliJ* mRNA encodes a protein participating in the cytoplasmic ATPase ring complex that is part of flagellar type III export apparatus and that coordinates flagellar protein export with assembly (Minamino *et al.*, 2000; Sajó *et al.*, 2014; Minamino *et al.*, 2017). FliJ acts as a chaperone, exporting flagellar proteins from the cytoplasm into the periplasm (Kang *et al.*, 2015). The Hsnc107 sRNA hybridizes with *fliJ* mRNA in the region of -10, upstream to the AUG, to position +6 (Additional data 4B).

The product of the *def* mRNA is the enzyme peptide deformylase (PDF). The PDF is an essential enzyme in bacteria responsible for the deformylation of the N-terminal methionine during protein synthesis, therefore, the promising target for antibacterial drugs (Chan *et al.*, 1997; Margolis *et al.*, 2000; Clements *et al.*, 2001, Ranjan *et al.*, 2016). The Hsnc107 sRNA hybridizes with the *def* mRNA in the same region detected in the *fliJ* mRNA (-10 to +6) (Additional data 4C).

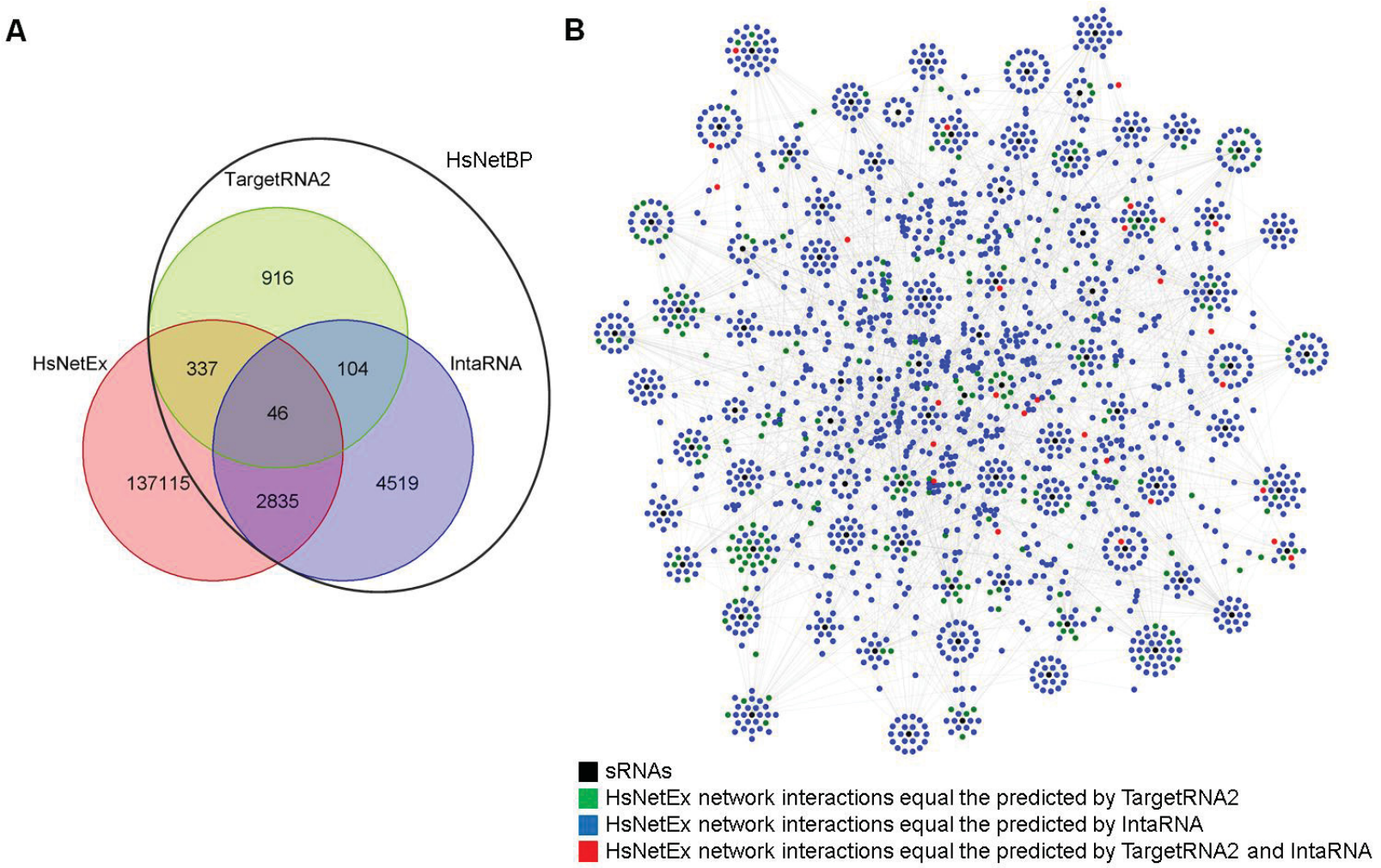


FIGURE 5. Number of sRNA-mRNAs interactions amongst HsNetEx, IntaRNA and TargetRNA2. (A) Venn diagram of interactions amongst HsNetEx, IntaRNA and TargetRNA2; (B) Graphical visualization of the 3,218 sRNAs-mRNA interactions of consensus Network converging between predicted by the IntaRNA and / or TargetRNA2 tools. The graphical visualization was done using Cytoscape 3.7.1 software.

We performed an enrichment analysis of functional categories using the COG database for the mRNAs of consensus network (Figure 6 and [2]Additional data 5). There is a large group of proteins that are involved in energy production and conversion (109), amino acid metabolism and transport (210), transcription (177), signal transduction mechanisms (161), carbohydrate transport and metabolism (121), cell wall structure and biogenesis (125) inorganic ion transport and metabolism (133) and general functional prediction (254) (Figure 7). However, many genes (469) do not present functional annotation and therefore are grouped in category S (no functional prediction).

Several sRNAs-mRNA interactions involving nitrogen metabolism mRNAs in of *H. seropedicae* SmR1 were observed, including *nifA (*Hsnc107*), nifH (*Hsnc010, Hsnc079 and Hsnc088*), nifQ (*Hsnc082)*, nifW (*Hsnc046 and Hsnc074*), nifU (*Hsnc096*) and nifS (*Hsnc070*)*, *fixI (*Hsnc003, Hsnc070, Hsnc079 and Hsnc106*), fixP (*Hsnc003*), fixO (*Hsnc015, Hsnc021and Hsnc092*), fixN (*Hsnc046), *fixG* (Hsnc055 and Hsnc072), *fixA* (Hsnc056), *fixU* (Hsnc077), *modC1* (Hsnc020, Hsnc074), *modB1* (Hsnc038, Hsnc058, Hsnc076 and Hsnc095), *fdx* (Hsnc037, Hsnc046, Hsnc061, Hsnc076 and Hsnc106), *fdxB* (Hsnc051), *fdxN* (Hsnc065).

[2] Additional data provided by email

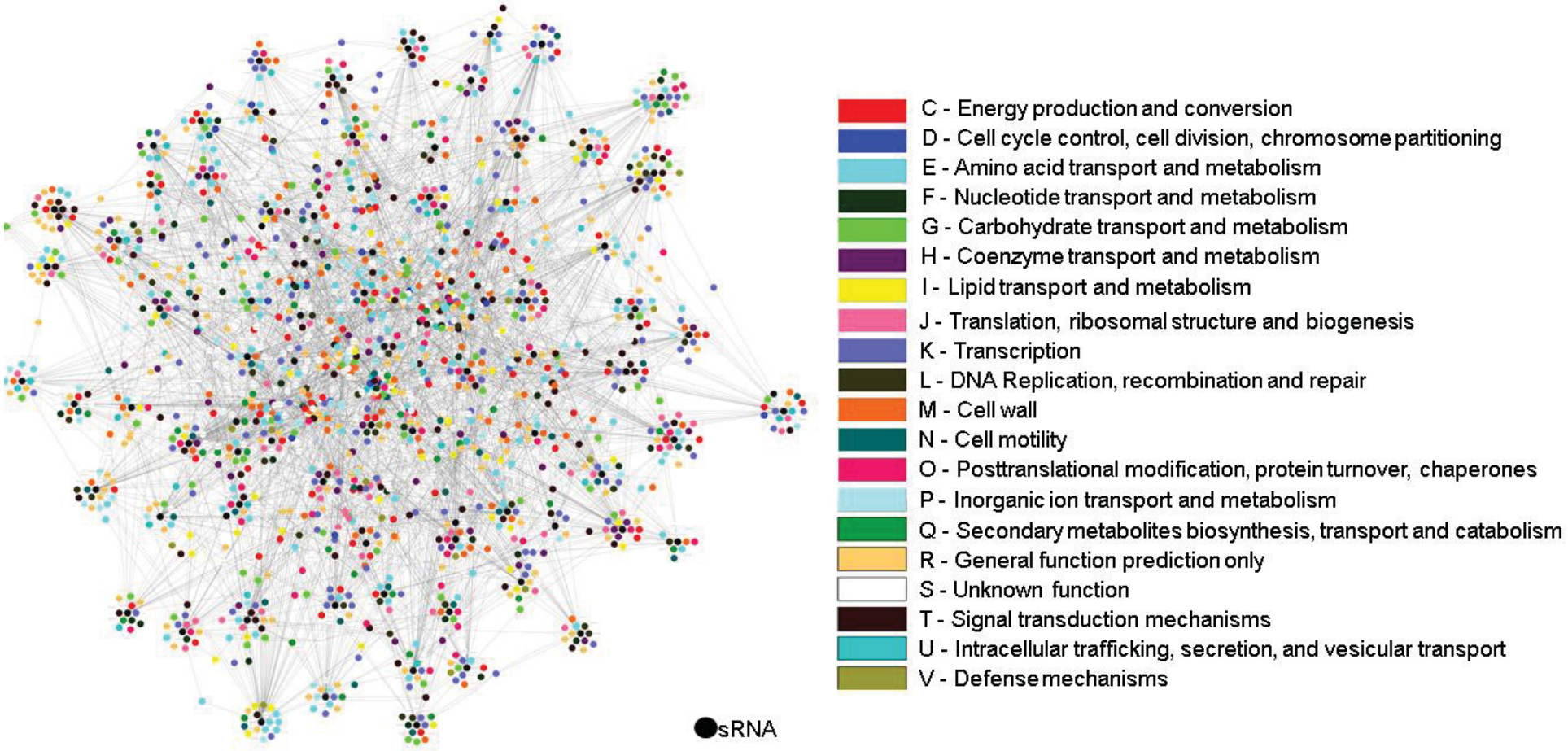


FIGURE 6: Consensus sRNA-mRNA Network for *H. seropedicae* SmR1. The mRNAs were labeled in different colors according to COG categories.

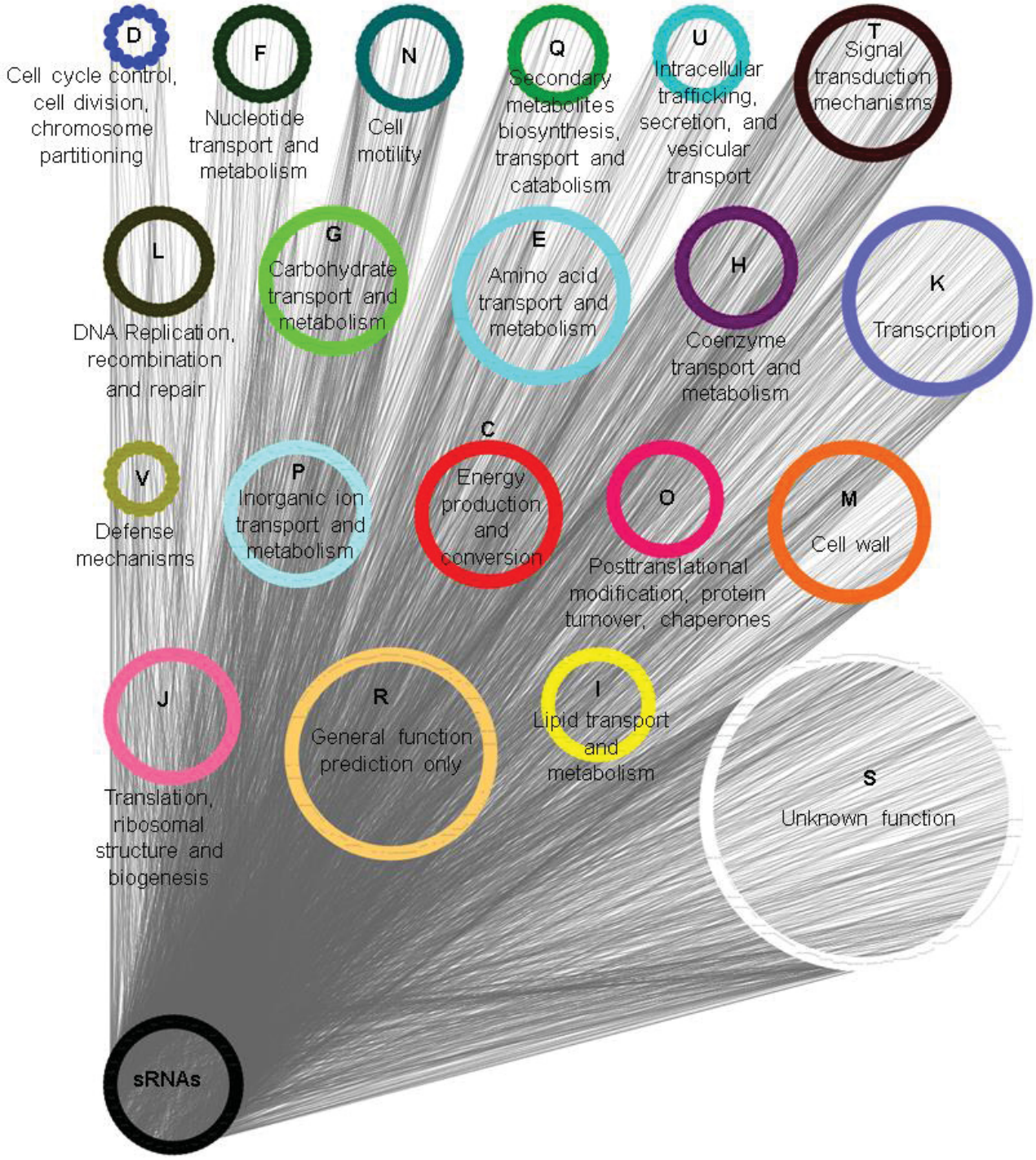


FIGURE 7: Network consensus of sRNAs-mRNAs of *H. seropedicae* SmR1 with genes grouped by function according to the COG database.

The distribution of the *H. seropedicae* SmR1 sRNA in functional groups according to the COG database is shown in the table 1. Most of the mRNAs that interact with sRNAs are grouped in the category R - General function prediction only, E - Amino acid transport and metabolism or T - Signal transduction mechanisms (Table 1). The Hsnc015 sRNA was the only one to do the largest amount of interactions (6) with N - Cell motility mRNAs category. The Hsnc007 presented 12 interactions with mRNAs grouped in the T - Signal transduction mechanisms category and 9 mRNAs related to K - Transcription category.

TABLE 1: Distribution of the *H. seropedicae* SmR1 sRNAs present in the consensus network that encode proteins categorized into functional groups according to the COG database.

sRNA	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
Hsnc003	4	4	5	1	1	1	1	2	6	0	2	3	0	4	1	7	5	6	3	1
Hsnc005	2	0	3	0	5	3	2	2	3	4	2	0	0	4	2	11	4	0	3	0
Hsnc007	3	0	2	1	3	0	4	1	9	1	6	5	2	4	3	9	3	12	2	0
Hsnc009	3	0	6	1	4	4	0	2	9	1	5	0	0	1	1	5	6	4	1	0
Hsnc010	2	1	5	0	6	2	2	1	1	1	2	1	0	5	0	6	1	3	2	0
Hsnc011	4	1	1	0	3	0	0	0	2	2	1	3	0	3	0	4	1	2	2	0
Hsnc013	1	0	1	0	0	2	0	0	10	0	7	0	1	2	0	2	5	3	2	0
Hsnc014	4	0	4	1	2	2	0	0	5	1	5	3	1	2	2	6	1	3	0	0
Hsnc015	2	0	3	0	1	1	1	2	3	2	5	6	1	3	0	4	3	6	3	0
Hsnc016	3	1	5	0	6	1	5	0	5	0	7	5	3	4	3	8	4	8	1	0
Hsnc017	1	0	3	0	3	0	2	1	3	0	1	2	1	2	1	5	8	4	2	0
Hsnc018	0	1	6	1	7	0	0	2	0	0	2	1	1	2	0	8	1	0	1	0
Hsnc019	2	1	5	0	4	1	2	2	2	2	3	2	1	1	0	10	7	4	0	0
Hsnc020	6	1	3	1	4	1	0	0	0	2	8	0	3	8	2	7	4	0	1	0
Hsnc021	5	1	5	3	6	1	1	0	2	1	5	3	1	4	1	5	4	4	2	0
Hsnc022	2	0	1	2	3	2	1	1	3	1	5	3	1	0	0	4	2	5	2	1
Hsnc025	2	2	2	1	1	3	2	2	4	2	0	2	0	1	2	4	5	5	0	0
Hsnc026	0	0	0	0	0	0	0	0	1	0	1	1	1	0	0	2	1	1	0	0
Hsnc027	1	0	9	3	4	2	3	4	7	1	2	2	1	3	1	6	7	7	1	1
Hsnc028	5	0	6	0	3	1	1	2	1	1	2	2	1	4	1	5	4	3	1	0

sRNA	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
Hsnc030	1	1	5	0	3	4	1	4	7	1	3	3	3	1	5	9	0	3	1	0
Hsnc034	2	0	2	2	2	3	1	2	6	2	6	1	3	5	3	6	5	3	4	1
Hsnc036	3	0	4	1	4	2	1	2	1	0	3	5	2	4	0	13	4	8	2	0
Hsnc037	5	0	1	0	1	0	1	1	2	1	3	0	0	0	0	3	1	1	0	0
Hsnc038	2	0	3	2	2	2	1	4	0	1	1	2	3	3	1	4	4	2	0	0
Hsnc039	0	0	2	0	2	0	2	0	1	0	2	1	1	1	0	2	2	2	1	0
Hsnc040	2	0	3	1	1	0	0	0	5	0	4	2	2	2	1	4	3	3	0	1
Hsnc041	0	0	2	1	0	1	2	1	2	4	3	4	1	0	1	4	0	5	1	0
Hsnc042	2	0	7	3	2	3	2	5	4	1	6	3	3	6	3	6	4	3	1	0
Hsnc043	1	0	5	1	5	3	3	2	0	2	4	1	0	3	2	7	5	4	0	1
Hsnc045	5	0	3	0	1	4	1	0	3	0	2	0	0	3	0	3	4	1	0	0
Hsnc046	3	0	5	2	2	3	3	0	0	0	3	2	4	6	1	7	5	5	1	1
Hsnc047	2	1	2	0	2	1	0	2	0	0	0	1	0	1	1	3	3	1	0	1
Hsnc048	1	1	2	3	2	0	1	0	2	0	6	0	2	2	1	5	4	2	0	0
Hsnc049	0	0	5	0	1	0	2	0	4	1	2	3	0	1	1	5	2	4	0	0
Hsnc051	4	0	4	0	3	1	1	1	2	1	1	3	0	4	1	6	4	2	2	1
Hsnc052	2	0	7	0	4	0	0	0	3	2	2	2	0	5	2	10	1	2	2	0
Hsnc053	4	0	5	2	7	2	2	2	7	3	2	4	2	3	3	7	7	2	2	0
Hsnc055	7	2	5	0	2	2	1	2	9	2	0	5	2	5	1	4	9	10	1	0
Hsnc056	3	0	6	1	2	2	3	6	4	1	3	0	1	4	0	6	4	1	0	0
Hsnc058	1	1	6	1	0	1	2	1	1	0	3	3	1	5	2	8	2	3	2	1
Hsnc059	4	1	6	1	4	1	1	3	4	1	3	2	1	2	0	6	3	2	1	1
Hsnc060	2	0	3	0	1	0	1	1	0	0	4	0	0	2	0	1	4	0	0	0
Hsnc061	2	0	4	1	3	1	1	0	3	3	2	1	1	0	1	6	3	2	0	0
Hsnc062	1	0	3	0	0	1	0	4	1	0	1	0	1	2	0	3	1	1	0	1
Hsnc063	0	0	3	1	1	2	1	3	3	1	2	0	0	3	1	1	4	1	0	0
Hsnc064	1	0	4	1	5	4	0	1	2	1	3	1	0	3	1	9	1	0	1	0
Hsnc065	4	0	8	0	4	2	3	1	2	2	0	1	2	4	1	6	3	2	1	0
Hsnc066	1	0	4	2	2	0	0	1	1	1	2	2	1	1	1	4	5	4	0	0
Hsnc070	1	0	7	1	5	5	4	3	3	1	5	0	2	5	2	5	7	2	1	1
Hsnc071	2	0	3	1	3	0	0	0	2	3	7	1	0	8	2	11	3	5	1	1
Hsnc072	4	2	8	0	4	2	3	1	6	4	5	5	0	1	1	8	9	10	1	1

sRNA	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
Hsnc073	0	1	6	1	4	0	2	1	4	1	0	5	1	3	1	8	3	7	1	1
Hsnc074	2	0	3	1	4	2	0	1	4	0	3	1	0	8	1	7	5	5	0	0
Hsnc076	4	0	2	1	0	2	1	0	1	2	2	2	1	6	1	2	3	2	1	0
Hsnc077	3	0	0	3	1	0	1	1	1	0	2	0	1	3	1	5	5	3	2	0
Hsnc079	3	2	8	2	3	1	4	0	8	2	7	1	6	6	1	9	7	9	0	3
Hsnc081	1	0	7	1	3	1	0	3	6	1	1	0	0	2	1	4	0	1	0	0
Hsnc082	1	1	5	1	0	1	0	1	3	1	2	3	1	3	1	5	3	3	0	0
Hsnc084	1	0	5	0	1	2	0	2	3	0	3	2	0	1	0	2	2	4	3	0
Hsnc086	2	1	6	0	2	1	1	2	4	0	5	1	2	1	1	6	1	4	2	0
Hsnc087	0	0	3	0	2	1	3	0	0	1	1	0	1	2	1	7	4	1	0	0
Hsnc088	0	0	2	1	1	2	3	2	2	2	1	3	0	2	3	9	7	5	2	2
Hsnc090	2	0	6	2	3	3	2	1	3	2	7	1	2	2	1	5	3	2	2	2
Hsnc092	3	1	5	0	2	0	0	3	6	0	3	3	0	2	1	4	3	2	3	0
Hsnc095	3	0	6	0	5	2	2	0	0	0	2	2	1	2	2	6	3	4	3	1
Hsnc096	2	0	8	0	2	1	4	1	6	0	1	1	1	3	1	6	3	7	1	0
Hsnc097	6	0	4	1	1	3	0	0	6	0	4	1	0	1	2	5	0	5	0	2
Hsnc099	2	1	4	0	2	7	1	7	4	0	3	3	0	3	2	7	4	3	1	0
Hsnc100	0	0	0	0	0	0	1	2	2	0	2	3	0	2	0	2	2	5	0	0
Hsnc101	0	1	4	0	0	1	4	0	2	2	2	5	2	5	1	5	5	6	2	0
Hsnc102	4	0	2	1	0	0	1	0	8	1	1	4	3	0	1	2	7	5	0	0
Hsnc104	1	0	3	0	2	0	1	0	0	1	3	0	1	2	2	5	2	1	0	0
Hsnc105	3	0	5	1	4	3	3	4	3	0	5	2	3	4	3	6	4	5	1	1
Hsnc106	4	1	7	2	3	1	0	1	2	1	1	3	2	6	0	4	2	6	0	0
Hsnc107	3	0	3	0	2	0	1	1	2	0	2	2	2	3	2	2	1	1	2	0
Hsnc108	3	0	9	2	3	3	7	1	6	2	3	2	1	6	4	7	7	3	2	1

The highest number of interactions, among categories, made by each sRNA is highlighted in orange. The number of interactions made by each sRNA with an mRNA without know function (S category) is highlighted in yellow. Letters represents the COG functions: C - Energy production and conversion; D - Cell cycle control, cell division, chromosome partitioning; E - Amino acid transport and metabolism; F - Nucleotide transport and metabolism; G - Carbohydrate transport and metabolism; H - Coenzyme transport and

metabolism; I - Lipid transport and metabolism; J - Translation, ribosomal structure and biogenesis; K – Transcription; L - DNA Replication, recombination and repair; M - Cell wall; N - Cell motility; O - Posttranslational modification, protein turnover, chaperones; P - Inorganic ion transport and metabolism; Q - Secondary metabolites biosynthesis, transport and catabolism; R - General function prediction only; S - Unknown function; T - Signal transduction mechanisms; U - Intracellular trafficking, secretion, and vesicular transport; V - Defense mechanisms.

5.5. DISCUSSION

The sRNAs are the key elements in the global network of post-transcriptional regulation in bacteria (Melamed *et al.*, 2016; Nitzan *et al.*, 2017; Brosse and Guillier, 2018). Most of the sRNAs found in bacteria are trans-encoded and interact with mRNAs by imperfect base pairing. This kind of sRNAs may have dozens of target mRNAs (Pain *et al.*, 2015). The functional characterization and target identification of sRNAs requires extensive experiments, but bioinformatics approaches can help in this process. In this work, we applied bioinformatics to investigate interactions between sRNAs and mRNAs in *H. seropedicae* SmR1. We applied two methodologies to infer a sRNA-mRNA network, the first one uses RNA expression data and the second one uses the sequence of sRNAs to search for base pairing regions on mRNAs sequences.

Five methods (Aracne, MRNET, CLR, Genie3 and DimReduction) to infer sRNA-mRNA interactions networks were applied and we choose the interactions that were common among the methods to create a robust network that presents more precise interactions. A similar methodology was used by Marbach *et al.* (2012) in the Dialogue on Reverse Engineering Assessment and Methods (DREAM) project that integrates the outputs from multiple inference methods to build a robust and high-performance network across diverse data sets. We could confirm that the performance of inference methods based on expression data varies greatly according to the study dataset (Lopes *et al.*, 2009; Marbach *et al.*, 2012) since the number of sRNA-mRNA interactions (output) was different and the MRNET and CLR presented the highest number of common sRNA-mRNA interactions.

The network inference methods used in this work have been employed in other case studies. Modi *et al.* (2011) using CLR and a compendium with expression data was able to predict functional roles and regulatory interactions for *E. coli* sRNAs. Basso *et al.* (2005)

used Aracne and was able to report the reconstruction of regulatory networks from expression profiles of human B cells. Meisig and Blüthgen *et al.* (2018) analyzed the algorithms Aracne, MRNET and CLR for inference the gene regulatory network underlying differentiation of embryonic stem cells. Aibar *et al.* (2017) used Genie3 to identify a set of genes that are coexpressed with transcription factors for inference from a single-cell regulatory network. Mendonça *et al.* (2013) used DimReduction to find interactions associated with clinical immunity to *Plasmodium vivax* malaria.

In the consensus network, we have identified several sRNAs that interact with mRNAs coding proteins involved in nitrogen fixation. Many of these mRNAs presented more than one sRNA as a post-transcriptional regulator. This kind of regulation is not surprising since it has already been demonstrated that an mRNA can be regulated by different sRNAs. This is the case of the CsgD, a transcriptional regulator in the biofilm formation, which is negatively regulated by at least six sRNAs that bind to partially overlapping regions in the 5'-UTR of *csgD* mRNA, with mechanistically different consequences (Mika and Hengge, 2014). Another example was demonstrated by De Lay and Gottesman (2012) who identified in *E. coli flhDCmRNA* is negatively regulated by six sRNAs, ArcZ, OmrA, OmrB, OxyS, SdsR and GadY, and positively regulated by one sRNA, McaS. Our results indicate that in *H. seropedicae* SmR1 several sRNAs may be regulating a single mRNA and this should be confirmed experimentally.

Through our network approach, we have been able to identify interesting relationships, such as the Hsnc020-htrB interaction, which presented fairly negative hybridization energy. Hsnc020 sRNA hybridizes with the mRNA in the position -1 to +20, being -1 upstream to AUG and next to RBS. We suggest that the sRNA-mRNA paring in this region may disrupt the perfect fit of the ribosome to the mRNA, causing instability of alignment between ribosome-mRNA and preventing translation of mRNA.

H. seropedicae presents four Betaproteobacteria Toxic Small RNA - tsRNA (Hsnc059, Hsnc063, Hsnc086 and Hsnc107). They are so called because they are unique to Betaproteobacteria and when expressed in *E. coli* inhibited their growth (Kimelman *et al.*, 2012; Sass, 2015). Targets of tsRNAs are not yet known, however, how these sRNAs exhibit a conserved motif complementary to the consensus ribosomal binding site (RBS) sequence, possibly interact with their targets by blocking the RBS in the 5 'UTR region of several genes (Kimelman *et al.*, 2012). In our consensus network the Hsnc107 sRNA can interact with the mRNA coding for the peptide deformylase, an essential enzyme involved in the protein

synthesis in bacteria (Chan *et al.*, 1997; Margolis *et al.*, 2000; Clements *et al.*, 2001, Ranjan *et al.*, 2016). We note that Hsnc107 possibly hybridizes in the RBS region of the *def*mRNA (-10, upstream to the AUG, to position +6). Since Hsnc107 showed interaction with the mRNA of a promising target for antibacterial drugs (Clements *et al.*, 2001; Sharma *et al.*, 2009), and as it is only known that tsRNAs inhibit bacterial growth, we suggest that *def*mRNA is promising for tsRNAs target research.

Another potential target of Hsnc107 is *fliJ*mRNA. Previously we have already observed that Hsnc107 shows a high expression in the presence of naringenin (Dobrzanski *et al.*, 2019 submitted) whereas f*liJ* gene expression was downregulated in the presence of naringenin (Tadra-Sfeir *et al.*, 2015). Here we observed that at the post-transcriptional level the *fliJ*mRNA can interacts with the Hsnc107 sRNA. This sRNA hybridizes to the RBS region of the *fliJ*mRNA (-10, upstream to the AUG, to position +6) and this suggest that *fliJ*mRNA translation can be inhibited since the ribosome cannot correctly bind to the mRNA.

Network inference based on expression data is an important approach to detect gene clusters that exhibit transcriptional responses across various cellular conditions. This method can be improved with the information of possible base pairings between the sRNAs and the mRNAs. So, we used TargetRNA2 and IntaRNA tools to search for possible regions of base pairing for target prediction of sRNAs in *H. seropedicae* SmR1 and we got the HsNetBP network. The HsNetBP network merged with HsNetEx network resulted in a consensus network (Figures 1 and 4) with more accurate results which can be more suitable for experimental validation.

5.6. CONCLUSIONS

We reported the inference of networks of sRNA-mRNA interaction in *H. seropedicae* SmR1 from two network inference approaches, using expression data and searching for base pairing regions with mRNAs. We were able to identify mRNAs with high potential to be targets of sRNAs, such as Hsnc107 previously identified as Betaproteobacteria Toxic Small RNA. The sRNAs of *H. seropedicae* SmR1 interact with several mRNAs involved in different metabolic pathways, including mRNAs that encode proteins for nitrogen fixation.

Additional data 1: **RNA -Seq data analysis**

Additional data 2: **sRNA-mRNA interactions inferred by methods based on expression data.** The common interactions found in at least 2 networks based on expression data were selected to compose the network called HsNetEx.

Additional data 3: **sRNA-mRNA common interactions between HsNetEx and the predictions of TargetRNA2 and IntaRNA.**

Additional data 4A and 4B: Base pairing between sRNA-mRNA. (A) Region of interaction between Hsnc020 sRNA and htrBmRNA. (B) Region of interaction between Hsnc107 sRNA and fliJmRNA; (C) Region of interaction between Hsnc107 sRNA and the defmRNA.

Additional data 5: sRNA-mRNA consensus network and COG categories

5.7. REFERENCES

Aibar, S., González-Blas, C. B., Moerman, T., Imrichova, H., Hulselmans, G., Rambow, F., *et al.* (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature methods* 14(11), 1083. doi:10.1038/nmeth.4463.

Andrade, J.M., Pobre, V., Matos, A.M., and Arraiano, C.M. (2012). The crucial role of PNPase in the degradation of small RNAs that are not associated with Hfq. *RNA* 18(4), 844-855. doi: 10.1261/rna.029413.111.

Babski, J., Maier, L. K., Heyer, R., Jaschinski, K., Prasse, D., Jäger, D., *et al.* (2014). Small regulatory RNAs in archaea. *RNA biology* 11(5), 484-493. doi: 10.4161/rna.28452.

Baldani, J.I., and Baldani, V.L.D. (2005). History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. *Anais da Academia Brasileira de Ciências* 77, 549-579.

Baldani, J.I., Pot, B., Kirchhof, G., Falsen, E., Baldani, V.L., Olivares, F.L., *et al.* (1996). Emended description of *Herbaspirillum*; inclusion of [*Pseudomonas*] rubrisubalbicans, a milk plant pathogen, as *Herbaspirillum rubrisubalbicans* comb. nov.; and classification of a group of clinical isolates (EF group 1) as *Herbaspirillum* species 3. *Int J Syst Bacteriol* 46(3), 802-810. doi: 10.1099/00207713-46-3-802.

Balsanelli, E., Tadra-Sfeir, M. Z., Faoro, H., Pankievicz, V. C., de Baura, V. A., Pedrosa, F. O., *et al.* (2016). Molecular adaptations of *Herbaspirillum seropedicae* during colonization of the maize rhizosphere. *Environmental microbiology* 18(8), 2343-2356. doi: 10.1111/1462-2920.12887.

Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. *Nature genetics* 37(4), 382. doi: 10.1038/ng1532.

Brosse, A., and Guillier, M. (2018). Bacterial Small RNAs in Mixed Regulatory Networks. *Microbiology spectrum* 6(3). doi:10.1128/microbiolspec.RWR-0014-2017.

Busch, A., Richter, A. S., and Backofen, R. (2008). IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. *Bioinformatics*, 24(24), 2849-2856. doi:10.1093/bioinformatics/btn544.

Carrier, M.C., Lalaouna, D., and Massé, E. (2018). Broadening the definition of bacterial small RNAs: Characteristics and mechanisms of action. *Annual review of microbiology* 72, 141-161. doi: 10.1146/annurev-micro-090817.

Chan, M. K., Gong, W., Rajagopalan, P. R., Hao, B., Tsai, C. M., and Pei, D. (1997). Crystal structure of the *Escherichia coli* peptide deformylase. *Biochemistry* 36(45), 13904-13909. doi: 10.1021/bi9711543.

Clements, J. M., Beckett, R. P., Brown, A., Catlin, G., Lobell, M., Palan, S., *et al.* (2001). Antibiotic activity and characterization of BB-3497, a novel peptide deformylase inhibitor. *Antimicrobial agents and chemotherapy* 45(2), 563-570. doi: 10.1128/AAC.45.2.563–570.2001.

De Lay, N., and Gottesman, S. (2012). A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. *Molecular microbiology* 86(3), 524-538. doi: 10.1111/j.1365-2958.2012.08209.x.

Dobrzanski T., Pobre V., Moreno L., Cirino H., Monteiro R. A., Pedrosa F. O., Souza E. M., Arraiano C. M., Steffens M. B. R. (2019). In silico prediction and expression profile analysis of small non-coding RNAs in *Herbaspirillum seropedicae* SmR1. *Frontiers in Microbiology* (submitted).

Dressaire, C., Pobre, V., Laguerre, S., Girbal, L., Arraiano, C.M., and Cocaign-Bousquet, M. (2018). PNPase is involved in the coordination of mRNA degradation and expression in stationary phase cells of *Escherichia coli*. *BMC genomics* 19(1), 848. doi: 10.1186/s12864-018-5259-8.

Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., *et al.* (2007). Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS biology* 5(1), e8. doi:10.1371/journal.pbio.0050008.

Gottesman, S., and Storz, G. (2011). Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harbor perspectives in biology* 3(12). doi: 10.1101/cshperspect.a003798.

Hoe, C.H., Raabe, C.A., Rozhdestvensky, T.S., and Tang, T.H. (2013). Bacterial sRNAs: regulation in stress. *International Journal of Medical Microbiology* 303(5), 217-229. doi: 10.1016/j.ijmm.2013.04.002.

Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PloS one* 5(9), e12776. doi:10.1371/journal.pone.0012776.

Jones, B. D., Nichols, W. A., Gibson, B. W., Sunshine, M. G., and Apicella, M. A. (1997). Study of the role of the htrB gene in *Salmonella typhimurium* virulence. *Infection and immunity* 65(11), 4778-4783. https://iai.asm.org/content/iai/65/11/4778.full.pdf.

Kang, X.M., Wang, F.F., Zhang, H., Zhang, Q., and Qian, W. (2015). Genome-wide identification of genes necessary for biofilm formation by nosocomial pathogen *Stenotrophomonas maltophilia* reveals that orphan response regulator FsnR is a critical modulator. *Appl. Environ. Microbiol* 81(4), 1200-1209. doi:10.1128/AEM.03408-14.

Batista, M. B., Sfeir, M. Z., Faoro, H., Wassem, R., Steffens, M. B., Pedrosa, F. O., *et al.* (2013). The *Herbaspirillum seropedicae* SmR1 Fnr orthologs controls the cytochrome composition of the electron transport chain. *Scientific reports* 3, 2544. doi: 10.1038/srep02544.

Kery, M. B., Feldman, M., Livny, J., and Tjaden, B. (2014). TargetRNA2: identifying targets of small regulatory RNAs in bacteria. *Nucleic acids research* 42(W1), W124-W129. doi: 10.1093/nar/gku317.

Kimelman, A., Levy, A., Sberro, H., Kidron, S., Leavitt, A., Amitai, G., *et al.* (2012). A vast collection of microbial genes that are toxic to bacteria. *Genome Research* 22(4), 802--809. doi: 10.1101/gr.133850.111.

Lalaouna, D., Simoneau-Roy, M., Lafontaine, D., and Massé, E. (2013). Regulatory RNAs and target mRNA decay in prokaryotes. *Biochimica Et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1829(6-7), 742-747. doi: 10.1016/j.bbagrm.2013.02.013.

Liu, J.M., and Camilli, A. (2010). A broadening world of bacterial small RNAs. *Current opinion in microbiology* 13(1), 18-23. doi: 10.1016/j.mib.2009.11.004

Lopes, F. M., Martins, D. C., and Cesar, R. M. (2009, May). Comparative study of GRNs inference methods based on feature selection by mutual information. *In 2009 IEEE International Workshop on Genomic Signal Processing and Statistics* (pp. 1-4). IEEE. doi: 10.1109/GENSIPS.2009.5174334.

Lopes, F.M., Martins, D.C., and Cesar, R.M. (2008). Feature selection environment for genomic applications. *BMC bioinformatics* 9(1), 451. doi: 10.1186/1471-2105-9-451.

Majdalani, Nadim; Vanderpool, Carin K.; Gottesman, Susan. Bacterial small RNA regulators. *Critical reviews in biochemistry and molecular biology* 2005, 40.2: 93-113. doi: 10.1080/10409230590918702.

Marbach, D., Costello, J. C., Küffner, R., Vega, N. M., Prill, R. J., Camacho, D. M., *et al.* (2012). Wisdom of crowds for robust gene network inference. *Nature methods* 9(8), 796. doi:10.1038/nmeth.2016.

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006, March). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In BMC bioinformatics (Vol. 7, No. 1, p. S7). *BioMed Central*. doi:10.1186/1471-2105-7-S1-S7.

Margolis, P. S., Hackbarth, C. J., Young, D. C., Wang, W., Chen, D., Yuan, Z., *et al.* (2000). Peptide deformylase in *Staphylococcus aureus*: resistance to inhibition is mediated by mutations in the formyltransferase gene. *Antimicrobial agents and chemotherapy* 44(7), 1825-1831. doi: 10.1128/AAC.44.7.1825-1831.2000.

Massé, E., Escorcia, F. E., and Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes and development*, 17(19), 2374-2383. doi:10.1101/gad.1127103

Meisig, J., and Blüthgen, N. (2018). The gene regulatory network of mESC differentiation: a benchmark for reverse engineering methods. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373(1750), 20170222. doi: 10.1098/rstb.2017.0222.

Melamed, S., Peer, A., Faigenbaum-Romm, R., Gatt, Y.E., Reiss, N., Bar, A., *et al.* (2016). Global mapping of small RNA-target interactions in bacteria. *Molecular cell* 63(5), 884-897. doi: 10.1016/j.molcel.2016.07.026.

Mendonça, V. R., Queiroz, A. T., Lopes, F. M., Andrade, B. B., and Barral-Netto, M. (2013). Networking the host immune response in *Plasmodium vivax* malaria. *Malaria journal* 12(1), 69. doi: 10.1186/1475-2875-12-69.

Meyer, P. E., Lafitte, F., and Bontempi, G. (2008). minet: AR/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC bioinformatics* 9(1), 461. doi:10.1186/1471-2105-9-461.

Mika, F., and Hengge, R. (2014). Small RNAs in the control of RpoS, CsgD, and biofilm architecture of *Escherichia coli*. *RNA biology* 11(5), 494-507. doi: 10.4161/rna.28867.

Minamino, T., and Macnab, R. M. (2000). FliH, a soluble component of the type III flagellar export apparatus of *Salmonella*, forms a complex with FliI and inhibits its ATPase activity. *Molecular microbiology* 37(6), 1494-1503. doi: 10.1046/j.1365-2958.2000.02106.x.

Minamino, T., Kinoshita, M., and Namba, K. (2017). Fuel of the bacterial flagellar type III protein export apparatus. *In The Bacterial Flagellum - Humana Press* 3-16. doi: 10.1007/978-1-4939-6927-2_1.

Modi, S. R., Camacho, D. M., Kohanski, M. A., Walker, G. C., and Collins, J. J. (2011). Functional characterization of bacterial sRNAs using a network biology approach. *Proceedings of the National Academy of Sciences* 108(37), 15522-15527. doi: 10.1073/pnas.1104318108.

Nitzan, M., Rehani, R., and Margalit, H. (2017). Integration of bacterial small RNAs in regulatory networks. *Annual review of biophysics* 46, 131-148. doi: 10.1146/annurev-biophys-070816-034058.

Pain, A., Ott, A., Amine, H., Rochat, T., Bouloc, P., and Gautheret, D. (2015). An assessment of bacterial small RNA target prediction programs. *RNA biology* 12(5), 509-513. doi: 10.1080/15476286.2015.1020269.

Pankievicz, V. C. S., Camilios-Neto, D., Bonato, P., Balsanelli, E., Tadra-Sfeir, M. Z., Faoro, H., *et al.* (2016). RNA-seq transcriptional profiling of *Herbaspirillum seropedicae* colonizing wheat (*Triticum aestivum*) roots. *Plant molecular biology* 90(6), 589-603. doi: 10.1007/s11103-016-0430-6.

Phongsisay, V., Perera, V. N., and Fry, B. N. (2007). Expression of the htrB gene is essential for responsiveness of *Salmonella typhimurium* and *Campylobacter jejuni* to harsh environments. *Microbiology* 153(1), 254-262. doi: 10.1099/mic.0.29230-0.

Ranjan, A., Mercier, E., Bhatt, A., and Wintermeyer, W. (2017). Signal recognition particle prevents N-terminal processing of bacterial membrane proteins. *Nature communications* 8, 15562. doi: 10.1038/ncomms15562.

Richards, G. R., and Vanderpool, C. K. (2011). Molecular call and response: the physiology of bacterial small RNAs. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1809(10), 525-531. doi: 10.1016/j.bbagrm.2011.07.013.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., *et al.* (2000). Artemis: sequence visualization and annotation. *Bioinformatics* 16(10), 944-945. doi: 10.1093/bioinformatics/16.10.944.

Sajó, R., Liliom, K., Muskotál, A., Klein, Á., Závodszky, P., Vonderviszt, F., and Dobó, J. (2014). Soluble components of the flagellar export apparatus, FliI, FliJ, and FliH, do not deliver flagellin, the major filament protein, from the cytosol to the export

gate. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 1843(11), 2414-2423. doi: 10.1016/j.bbamcr.2014.07.004.

Saramago, M., Bárria, C., dos Santos, R.F., Silva, I.J., Pobre, V., Domingues, S., *et al.* (2014). The role of RNases in the regulation of small RNAs. *Current opinion in microbiology* 18, 105-115. doi: 10.1016/j.mib.2014.02.009.

Sass, A.M.a. (2015). Genome-wide transcription start site profiling in biofilm-grown Burkholderia cenocepacia J2315. *BMC Genomics* 16(1), 1-15. doi: 10.1186/s12864-015-1993-3.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., *et al.* (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13(11), 2498-2504. doi/10.1101/gr.1239303.

Sharma, A., Khuller, G. K., and Sharma, S. (2009). Peptide deformylase – a promising therapeutic target for tuberculosis and antibacterial drug discovery. *Expert Opinion on Therapeutic Targets* 13(7), 753–765. doi:10.1517/14728220903005590.

Silva, I.J., Barahona, S., Eyraud, A., Lalaouna, D., Figueroa-Bossi, N., Massé, E., and Arraiano, C.M. (2019). SraL sRNA interaction regulates the terminator by preventing premature transcription termination of rho mRNA. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1811589116.

Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell* 43(6), 880-891. doi: 10.1016/j.molcel.2011.08.022.

Subramanian, D., Bhasuran, B., and Natarajan, J. (2018). Genomic analysis of RNA-Seq and sRNA-Seq data identifies potential regulatory sRNAs and their functional roles in *Staphylococcus aureus*. *Genomics*. doi: 10.1016/j.ygeno.2018.09.016.

Tadra-Sfeir, M. Z., Faoro, H., Camilios-Neto, D., Brusamarello-Santos, L., Balsanelli, E., Weiss, V., *et al.* (2015). Genome wide transcriptional profiling of *Herbaspirillum seropedicae* SmR1 grown in the presence of naringenin. *Frontiers in microbiology* 6, 491. doi: 10.3389/fmicb.2015.00491

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic acids research* 28(1), 33-36. doi: 10.1093/nar/28.1.33.

Valverde, C., Livny, J., Schluter, J.P., Reinkensmeier, J., Becker, A., and Parisi, G. (2008). Prediction of *Sinorhizobium meliloti* sRNA genes and experimental detection in strain 2011. *BMC Genomics* 9**,** 416. doi: 10.1186/1471-2164-9-416.

Vogel, J., and Wagner, E.G.H. (2007). Target identification of small noncoding RNAs in bacteria. *Current opinion in microbiology* 10(3), 262-270. doi: 10.1016/j.mib.2007.06.001.

Waters, L.S., and Storz, G. (2009). Regulatory RNAs in bacteria. *Cell* 136(4), 615-628. doi: S0092-8674(09)00125-1.

Zhan, S., Dong, Y., Zhao, W., Guo, J., Zhong, T., Wang, L., *et al.* (2016). Genome-wide identification and characterization of long non-coding RNAs in developmental skeletal muscle of fetal goat. *BMC genomics* 17(1), 666. doi: 10.1186/s12864-016-3009-3.

5.8. ADDITIONAL DATA

5.8.1. Additional data 1: RNA -Seq data analysis

The RNA-Seq data used in this work are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-3435, E-MTAB-3646, E-MTAB-5741 e E-MTAB-3026. The eleven treatments used for Seq RNA analysis were:

(i) Control condition (CRT), bacteria grown on NFbHPN medium, using malate as a source of carbon (Tadra-Sfeir *et al.* 2015);
(ii) Naringenin condition (NAR), bacteria grown in NFbHPN medium, using malate as carbon source and in the presence of flavonoid naringenin (Tadra-Sfeir *et al.* 2015);
(iii) Nitrate condition (NIT), bacteria grown in NFbHPN medium, using malate as carbon source and nitrate as nitrogen source (Bonato *et al.* 2016);
(iv) High oxygen condition (ATO), bacteria grown in NfbHPN, medium using malate as carbon source in the presence of high amounts of oxygen (Batista *et al.*, 2013);
(v) Low oxygen condition (BXO), bacteria grown in NFbHPN medium, using malate as carbon source in the presence of low oxygen (Batista *et al.*, 2013);
(vi) Condition 1 day of root adhesion to maize (ADU), bacteria adhered to maize root surface after 1 inoculation (Balsanelli *et al.*, 2016);
(vii) Condition 3 days of root adhesion to maize (ADT), bacteria adhered to maize root surface after 3 days of inoculation (Balsanelli *et al.*, 2016);
(viii) Condition of maize planktonic bacteria recovered from the NFbHPN culture medium after one day of inoculation in maize root (PSU) (Balsanelli *et al.*, 2016);
(ix) Condition of maize planktonic bacteria recovered from the NFbHPN culture medium after three days of inoculation in maize root (PST) (Balsanelli *et al.*, 2016);
(x) Condition planktonic bacteria recovered after 3 days of wheat root inoculation (PLT) (Pankievicz *et al.*, 2016);
(xi) Condition bacteria adhered to wheat root surface after three days of inoculation (WRU) (Pankievicz *et al.*, 2016);

5.8.2. Additional data 3: sRNA-mRNA common interactions between HsNetEx and the predictions of TargetRNA2 and IntaRNA.

sRNA	mRNA	[3]Energy IntaRNA	[4]Energy targetRNA	Gene name	sRNA start	sRNA stop	mRNA start	mRN A stop	Function
Hsnc010	HSERO_RS15545	-39.3261	-10,44	-	4	24	-68	-48	sulfite exporter TauE/SafE family protein
Hsnc014	HSERO_RS20820	-18.2385	-9,72	-	158	174	-17	-2	ABC transporter ATP-binding protein
Hsnc017	HSERO_RS11260	-17.202	-12,54	-	1	16	-42	-28	aldehyde dehydrogenase
Hsnc020	HSERO_RS03655	-24.4382	-15,37	-	59	73	8	20	ABC transporter substrate-binding protein
Hsnc020	HSERO_RS18455	-30.1061	-10,84	gph	42	58	-3	14	phosphoglycolate phosphatase
Hsnc020	HSERO_RS21445	-236.397	-23,58	htrB	70	90	-1	20	lipid A biosynthesis acyltransferase
Hsnc020	HSERO_RS21590	-27.3883	-20,32	rep	50	67	-39	-23	ATP-dependent DNA helicase Rep
Hsnc030	HSERO_RS02880	-10.2055	-8,44	-	69	86	6	20	class I SAM-dependent methyltransferase
Hsnc041	HSERO_RS18980	-13.5699	-13,01	-	26	40	-21	-7	Derived by automated computational analysis using gene prediction method: Protein Homology.
Hsnc042	HSERO_RS02145	-13.1263	-11,06	dps	32	41	12	20	DNA starvation/stationary phase protection protein
Hsnc042	HSERO_RS17105	-17.6878	-12,06	dnaQ	7	21	-52	-40	DNA polymerase III subunit epsilon
Hsnc043	HSERO_RS06465	-12.4581	-12,68	-	1	9	-49	-40	OmpA family protein
Hsnc043	HSERO_RS07615	-11.4801	-10,14	-	2	10	11	19	Derived by automated computational analysis using gene prediction method: Protein Homology.
Hsnc043	HSERO_RS22765	-10.7701	-9,11	glmS	4	15	-35	-24	SIS domain-containing protein
Hsnc045	HSERO_RS03620	-22.1697	-8,33	-	46	62	-68	-52	LysR family transcriptional regulator
Hsnc046	HSERO_RS04585	-10.0494	-12,86	-	89	105	-23	-7	Derived by automated computational analysis using gene prediction method: Protein Homology.
Hsnc048	HSERO_RS15870	-19.2107	-11,26	-	51	70	-33	-15	metal ABC transporter substrate-binding protein
Hsnc048	HSERO_RS22795	-16.9976	-14,5	-	18	32	-18	-3	N-acetyltransferase
Hsnc049	HSERO_RS13155	-13.9158	-10,81	-	19	36	-36	-19	zinc ribbon domain-containing protein

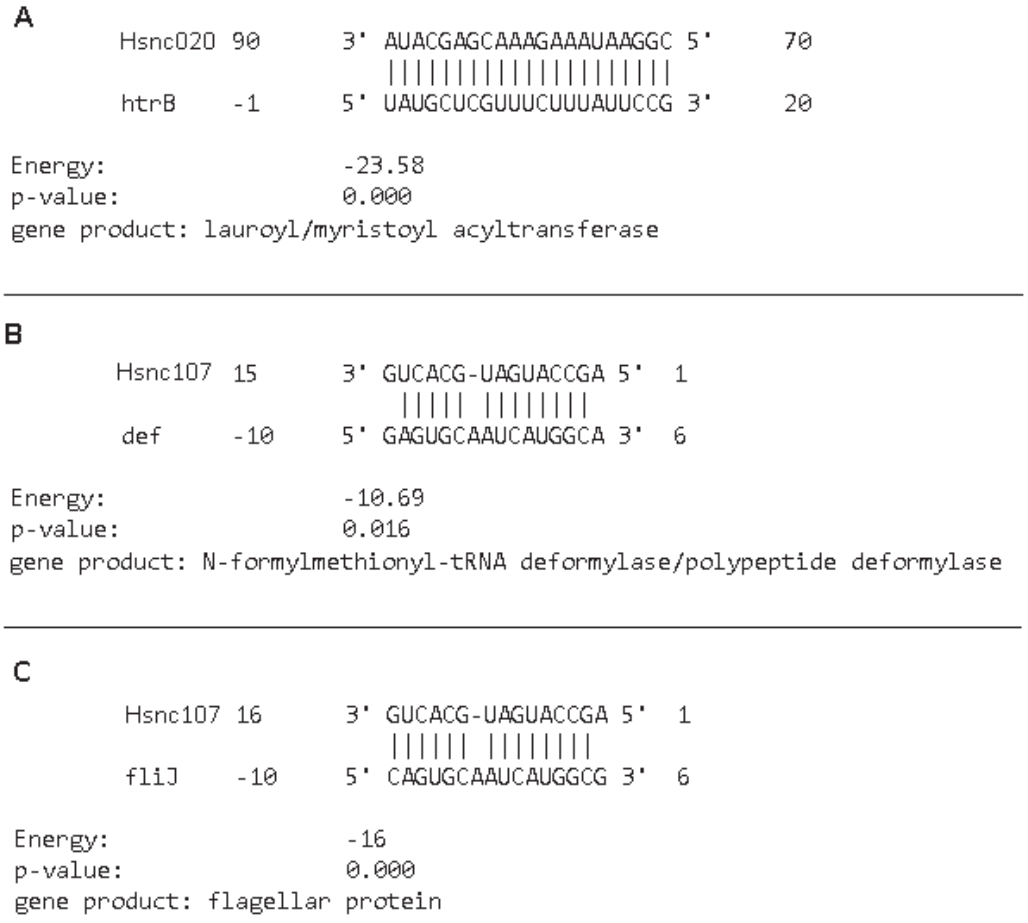
[3] The energy score (kcal/mol) of a predicted RNA-RNA interaction is the sum of the hybridization free energy of the interacting subsequences, and the free energies required to unfold the interaction sites in both RNA molecules.
[4] Thermodynamic energy (kcal/mol) of hybridization between the two RNA molecules.

sRNA	mRNA	[3]Energy IntaRNA	[4]Energy targetRNA	Gene name	sRNA start	sRNA stop	mRNA start	mRNA stop	Function
Hsnc051	HSERO_RS02290	-12.414	-10,91	pcbC	40	54	-5	10	isopenicillin N synthase family oxygenase
Hsnc052	HSERO_RS22115	-22.4755	-13,52	-	84	99	-22	-7	glycosyl transferase
Hsnc053	HSERO_RS09045	-13.0957	-9,53	tdcF	6	21	-3	13	RidA family protein
Hsnc053	HSERO_RS10445	-14.4211	-9,36	arnT	8	24	-64	-46	glycosyl transferase
Hsnc053	HSERO_RS19485	-14.8307	-15,39	ubiE	12	26	-74	-61	bifunctional demethylmenaquinone methyltransferase/2-methoxy-6-polyprenyl-14-benzoquinol methylase UbiE
Hsnc055	HSERO_RS02255	-12.8685	-11,71	-	4	20	3	18	TIGR00730 family Rossman fold protein
Hsnc055	HSERO_RS10455	-16.3055	-13,81	aspC	1	16	-16	-1	pyridoxal phosphate-dependent aminotransferase
Hsnc055	HSERO_RS17035	-17.9298	-8,6	-	1	16	-53	-39	Derived by automated computational analysis using gene prediction method: Protein Homology.
Hsnc055	HSERO_RS17380	-16.1696	-13,85	-	1	16	-9	7	peptidase
Hsnc055	HSERO_RS17605	-15.8585	-9,91	-	5	19	-19	-5	DUF1302 domain-containing protein
Hsnc055	HSERO_RS20155	-13.7628	-8,48	-	1	20	-22	-4	type VI secretion system baseplate subunit TssG
Hsnc055	HSERO_RS20600	-12.6328	-14,78	acrR	1	15	-16	-2	TetR/AcrR family transcriptional regulator
Hsnc055	HSERO_RS24315	-12.8021	-11,72	-	5	15	-15	-5	Derived by automated computational analysis using gene prediction method: Protein Homology.
Hsnc065	HSERO_RS05775	-22.4012	-10,86	-	86	100	-41	-27	LysR family transcriptional regulator
Hsnc072	HSERO_RS10540	-16.8413	-10,28	crcB	149	162	-14	1	fluoride efflux transporter CrcB
Hsnc092	HSERO_RS01495	-9.9209	-8,52	-	1	14	-6	9	B12-binding domain-containing radical SAM protein
Hsnc095	HSERO_RS22950	-21.1707	-11,48	entB	1	20	1	20	cysteine hydrolase
Hsnc102	HSERO_RS21960	-15.0871	-9,42	-	31	45	-14	1	gamma-glutamyl-gamma-aminobutyrate hydrolase
Hsnc104	HSERO_RS09965	-14.9567	-9,49	epsM	135	149	-28	-13	cholera toxin secretion EpsM protein
Hsnc105	HSERO_RS04455	-12.0076	-9,21	-	23	40	-23	-6	major tail tube protein
Hsnc105	HSERO_RS17380	-13.5012	-8,4	-	67	82	6	20	peptidase
Hsnc106	HSERO_RS10540	-31.3129	-12,39	crcB	42	57	-12	4	fluoride efflux transporter CrcB
Hsnc106	HSERO_RS22675	-15.8484	-8,53	tsr	47	58	-13	-2	methyl-accepting chemotaxis protein
Hsnc106	HSERO_RS24435	-18.1603	-14,93	-	81	92	-14	-3	Derived by automated computational analysis using gene prediction method: Protein Homology.

sRNA	mRNA	[3]Energy IntaRNA	[4]Energy targetRNA	Gene name	sRNA start	sRNA stop	mRNA start	mRN A stop	Function
Hsnc107	HSERO_RS10290	-16.7193	-16	fliJ	1	16	-10	6	flagellar export protein FliJ
Hsnc107	HSERO_RS10835	-17.9344	-10,69	def	1	15	-10	6	peptide deformylase
Hsnc107	HSERO_RS21240	-19.2246	-9,4	-	1	12	-7	7	amino acid ABC transporter permease

5.8.3. Additional data 4: Base pairing between sRNA-mRNA.

(A) Region of interaction between Hsnc020 sRNA and *htrB*mRNA; (B) Region of interaction between Hsnc107 sRNA and *fliJ*mRNA; (C) Region of interaction between Hsnc107 sRNA and the *def*mRNA.



6. CONCLUSÕES GERAIS

1. *H. seropedicae* SmR1 expressa 117 sRNAs classificados como *riboswitches* (10), codificados *in cis* (26) e codificados *in trans* (81).
2. A maior parte dos sRNAs de *H. seropedicae* SmR1 não estão descritos na literatura. Apenas 20 sRNAs apresentam identidade definida, incluindo os sRNAs *housekeeping* ssrS or 6S RNA (Hsnc050), 4.5S RNA (Hsnc001) e tmRNA (Hsnc083).
3. As condições ambientais e nutricionais afetam a expressão de sRNAs. Uma quantidade maior de sRNAs (62) são mais expressos na condição nitrato que nas condições controle e naringenina. Enquanto que outros 24 sRNAs são mais expressos na condição de naringenina do que nas condições de controle e nitrato.
4. A expressão de cinco sRNAs (Hsnc050, Hsnc028, Hsnc042, Hsnc073 e Hsnc082) foi confirmada por northern blot. Estes sRNAs são expressos nas condições controle, nitrato e presença de naringenina bem como na fase de crescimento exponencial e estacionária. Os sRNAs Hsnc073 e Hsnc082 apresentam um tamanho menor do que era predito por bioinformática.
5. A rede consenso apresenta interações sRNA-mRNA com alto potencial para serem testadas em laboratório. É o caso de Hsnc020-*htrB*, Hsnc107-*fliJ* e Hsnc107-*def*.
6. Os sRNAs de *H. seropedicae* SmR1 estão interagindo com mRNAs que participam de diversas funções metabólicas, incluindo a fixação de nitrogênio. Entretanto, a maioria dos mRNAs que interagem com sRNAs estão agrupados na categoria R-Predição de função geral apenas, E – Metabolismo e transporte de amino ácidos ou T – Mecanismos de transdução de sinal.

7. REFERENCES

AIBAR, S.; GONZÁLEZ-BLAS, C. B.; MOERMAN, T.; et al. SCENIC: Single-cell regulatory network inference and clustering. **Nature Methods**, v. 14, n. 11, p. 1083–1086, 2017.

ALTUVIA, S. Identification of bacterial small non-coding RNAs: experimental approaches. **Current opinion in microbiology**, v. 10, n. 3, p. 257–261, 2007.

ALVES, G. C.; VIDEIRA, S. S.; URQUIAGA, S.; REIS, V. M. Differential plant growth promotion and nitrogen fixation in two genotypes of maize by several *Herbaspirillum* inoculants. **Plant and Soil**, v. 387, n. 1–2, p. 307–321, 2014.

ARRAIANO, C. M.; ANDRADE, J. M.; DOMINGUES, S.; et al. The critical role of RNA processing and degradation in the control of gene expression. **FEMS Microbiology Reviews**, v. 34, n. 5, p. 883–923, 2010.

BABICKI, S.; ARNDT, D.; MARCU, A.; et al. Heatmapper: web-enabled heat mapping for all. **Nucleic acids research**, v. 44, n. W1, p. W147–W153, 2016.

BABSKI, J.; MAIER, L. K.; HEYER, R.; et al. Small regulatory RNAs in archaea. **RNA Biology**, v. 11, n. 5, 2014.

BAK, G.; LEE, J.; SUK, S.; et al. Identification of novel sRNAs involved in biofilm formation, motility, and fimbriae formation in *Escherichia coli*. **Scientific Reports**, v. 5, n. October, p. 1–19, 2015.

BALACHANDAR, D.; SANDHIYA, G. S.; SUGITHA, T. C. K.; KUMAR, K. Flavonoids and growth hormones influence endophytic colonization and in planta nitrogen fixation by a diazotrophic Serratia sp. in rice. **World Journal of Microbiology and Biotechnology**, v. 22, n. 7, p. 707–712, 2006.

BALDANI, J. I.; BALDANI, V. L. D. History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. **Anais da Academia Brasileira de Ciências**, v. 77, n. 3, p. 549–579, 2005. SciELO Brasil.

BALDANI, J. I.; POT, B.; KIRCHHOF, G.; et al. Emended Description of *Herbaspirillum*; Inccusion of [*Pseudomonas*] *rubrisubalbicans*, a Mild Plant Pathogen, as *Herbaspirillum rubrisubalbicans* comb. nov.; and Classification of a Group of Clinical Isolates (EF Group 1) as *Herbaspirillum* Species 3. **International Journal of Systematic Bacteriology**, v. 46, n. 3, p. 802–810, 1996.

BARBOSA, H. C. DE S. Re-anotação do genoma de *Herbaspirillum seropedicae* SmR1 com dados de transcriptoma por RNA_Seq. , 2014.

BASSO, K.; MARGOLIN, A. A.; STOLOVITZKY, G.; et al. Reverse engineering of regulatory networks in human B cells. **Nature Genetics**, v. 37, n. 4, p. 382–390, 2005.

BASTIÁN, F.; COHEN, A.; PICCOLI, P.; et al. Production of indole-3-acetic acid and gibberellins A1and A3by Acetobacter diazotrophicus and *Herbaspirillum seropedicae* in chemically-defined culture media. **Plant Growth Regulation**, v. 24, n. 1, p. 7–11, 1998.

BATISTA, M. B.; TEIXEIRA, C. S.; SFEIR, M. Z. T.; et al. PHB biosynthesis counteracts redox stress in *Herbaspirillum seropedicae*. **Frontiers in microbiology**, v. 9, p. 472, 2018. Frontiers.

BATTITI, R. Using mutual information for selecting features in supervised neural net learning. **IEEE Transactions on neural networks**, v. 5, n. 4, p. 537–550, 1994. IEEE.

BEISEL, C. L.; STORZ, G. Base pairing small RNAs and their roles in global regulatory networks. **FEMS Microbiology Reviews**, v. 34, n. 5, p. 866–882, 2010.

BOBROVSKYY, M.; VANDERPOOL, C. K. Regulation of Bacterial Metabolism by Small RNAs Using Diverse Mechanisms. **Annual Review of Genetics**, v. 47, n. 1, p. 209–232, 2013.

BONATO, P.; BATISTA, M. B.; CAMILIOS-NETO, D.; et al. RNA-seq analyses reveal insights into the function of respiratory nitrate reductase of the diazotroph *Herbaspirillum seropedicae*. **Environmental microbiology**, v. 18, n. 8, p. 2677–2688, 2016.

BRANTL, S. Regulatory mechanisms employed by cis-encoded antisense RNAs. **Current Opinion in Microbiology**, v. 10, n. 2, p. 102–109, 2007.

BREAKER, R. R. Riboswitches and translation control. **Cold Spring Harbor Perspectives in Biology**, v. 10, n. 11, 2018.

BROSSE, A.; GUILLIER, M. Bacterial Small RNAs in Mixed Regulatory Networks. **Microbiology Spectrum**, v. 6, n. 3, 2018.

BROWNLEE, G. G. Sequence of 6S RNA of E. coli. **Nature New Biology**, v. 229, n. 5, p. 147, 1971. Nature Publishing Group.

BUSCH, A.; RICHTER, A. S.; BACKOFEN, R. IntaRNA: Efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. **Bioinformatics**, v. 24, n. 24, p. 2849–2856, 2008.

CALDELARI, I.; CHAO, Y.; ROMBY, P.; VOGEL, J. RNA-mediated regulation in pathogenic bacteria. **Cold Spring Harbor perspectives in medicine**, v. 3, n. 9, p. a010298, 2013. Cold Spring Harbor Laboratory Press.

CARDOSO, R. L. A. Análise genômica comparativa de bactérias do gênero *Herbaspirillum*. 2015.

CARRIER, M.-C.; LALAOUNA, D.; MASSÉ, E. Broadening the Definition of Bacterial Small RNAs: Characteristics and Mechanisms of Action. **Annual Review of Microbiology**, v. 72, n. 1, p. 141–161, 2018.

CHAN, M. K.; GONG, W.; RAJAGOPALAN, P. T. R.; et al. Crystal structure of the *Escherichia coli* peptide deformylase. **Biochemistry**, v. 36, n. 45, p. 13904–13909, 1997.

CHAO, Y.; LI, L.; GIRODAT, D.; et al. In vivo cleavage map illuminates the central role of RNase E in coding and non-coding RNA pathways. **Molecular cell**, v. 65, n. 1, p. 39–51, 2017.

CHUBATSU, L. S.; MONTEIRO, R. A.; DE SOUZA, E. M.; et al. Nitrogen fixation control in Herbaspirillum seropedicae. **Plant and Soil**, v. 356, n. 1–2, p. 197–207, 2012.

CLEMENTS, J. M.; BECKETT, R. P.; BROWN, A.; et al. Antibiotic activity and characterization of BB-3497, a novel peptide deformylase inhibitor. **Antimicrobial Agents and Chemotherapy**, v. 45, n. 2, p. 563–570, 2001.

COENYE, T.; DREVINEK, P.; MAHENTHIRALINGAM, E.; et al. Identification of putative noncoding RNA genes in the *Burkholderia cenocepacia* J2315 genome. **FEMS Microbiology Letters**, v. 276, n. 1, p. 83–92, 2007.

COLEMAN, J.; GREEN, P. J.; INOUYE, M. The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. **Cell**, v. 37, n. 2, p. 429–436, 1984.

CRUZ, L. M.; DE SOUZA, E. M.; WEBER, O. B.; et al. 16S ribosomal DNA characterization of nitrogen-fixing bacteria isolated from banana (*Musa* spp.) and pineapple (*Ananas comosus* (L.) Merril). **Appl. Environ. Microbiol.**, v. 67, n. 5, p. 2375–2379, 2001. Am Soc Microbiol.

DAMBACH, M.; SANDOVAL, M.; UPDEGROVE, T. B.; et al. The Ubiquitous yybP-ykoY Riboswitch Is a Manganese-Responsive Regulatory Element. **Molecular Cell**, v. 57, n. 6, p. 1099–1109, 2015.

DELIHAS, N. Discovery and characterization of the first non-coding RNA that regulates gene expression, micF RNA: A historical perspective . **World Journal of Biological Chemistry**, v. 6, n. 4, p. 272, 2015.

DELIHAS, N.; FORST, S. MicF: An antisense RNA gene involved in response of *Escherichia coli* to global stress factors. **Journal of Molecular Biology**, v. 313, n. 1, p. 1–12, 2001.

DENNIS, P. P.; OMER, A. Small non-coding RNAs in Archaea. **Current Opinion in Microbiology**, v. 8, n. 6, p. 685–694, 2005.

DERMAN, Y.; SÖDERHOLM, H.; LINDSTRÖM, M.; KORKEALA, H. Role of csp genes in NaCl, pH, and ethanol stress response and motility in *Clostridium botulinum* ATCC 3502. **Food Microbiology**, v. 46, p. 463–470, 2015.

Dobrzanski T., Pobre V., Moreno L., Cirino H., Monteiro R. A., Pedrosa F. O., Souza E. M., Arraiano C. M., Steffens M. B. R. *In silico* prediction and expression profile analysis of small non-coding RNAs in *Herbaspirillum seropedicae* SmR1. **Frontiers in Microbiology,** 2019 (submitted).

DRESSAIRE, C.; POBRE, V.; LAGUERRE, S.; et al. PNPase is involved in the coordination of mRNA degradation and expression in stationary phase cells of *Escherichia coli.* **BMC Genomics**, v. 19, n. 1, p. 1–13, 2018. BMC Genomics.

DUCKER, G. S.; RABINOWITZ, J. D. ZMP: A master regulator of one-carbon metabolism. **Molecular Cell**, v. 57, n. 2, p. 203–204, 2015.

DUTTA, T.; SRIVASTAVA, S. Small RNA-mediated regulation in bacteria: A growing palette of diverse mechanisms. **Gene**, v. 656, n. November 2017, p. 60–72, 2018.

EDWARDS, A. L.; REYES, F. E.; HÉROUX, A.; BATEY, R. T. Structural basis for recognition of S-adenosylhomocysteine by riboswitches. **Rna**, v. 16, n. 11, p. 2144–2155,

2010.

ESTRADA, G. A.; BALDANI, V. L. D.; DE OLIVEIRA, D. M.; URQUIAGA, S.; BALDANI, J. I. Selection of phosphate-solubilizing diazotrophic *Herbaspirillum* and *Burkholderia* strains and their effect on rice crop yield and nutrient uptake. **Plant and Soil**, v. 369, n. 1–2, p. 115–129, 2013.

FAITH, J. J.; HAYETE, B.; THADEN, J. T.; et al. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. **PLoS Biology**, v. 5, n. 1, p. 0054–0066, 2007.

FANER, M. A.; FEIG, A. L. Identifying and characterizing Hfq-RNA interactions. **Methods**, v. 63, n. 2, p. 144–159, 2013.

FUENTES, D. N.; CALDERÓN, P. F.; ACUÑA, L. G.; et al. Motility modulation by the small non-coding RNA SroC in *Salmonella Typhimurium*. **FEMS Microbiology Letters**, v. 362, n. 17, p. 1–8, 2015.

GHOSH, SUPARNA; DUREJA, C.; KHATRI, I.; et al. Identification of novel small RNAs in *Burkholderia cenocepacia* KC-01 expressed under iron limitation and oxidative stress conditions. **Microbiology**, v. 163, n. 12, p. 1924–1936, 2017. Microbiology Society.

GÓMEZ-LOZANO, M.; MARVIG, R. L.; MOLIN, S.; LONG, K. S. Genome-wide identification of novel small RNAs in *Pseudomonas aeruginosa*. **Environmental microbiology**, v. 14, n. 8, p. 2006–2016, 2012. Wiley Online Library.

GOUGH, C.; GALERA, C.; VASSE, J.; et al. Specific flavonoids promote intercellular root colonization of *Arabidopsis thaliana* by *Azorhizobium caulinodans* ORS571. **Molecular plant-microbe interactions**, v. 10, n. 5, p. 560–570, 1997. Am Phytopath Society.

GREENLEE, E. B.; STAV, S.; ATILHO, R. M.; et al. Challenges of ligand identification for the second wave of orphan riboswitch candidates. **RNA Biology**, v. 6286, n. November, p. 1–14, 2018.

GYANESHWAR, P.; JAMES, E. K.; REDDY, P. M.; LADHA, J. K. *Herbaspirillum* colonization increases growth and nitrogen accumulation in aluminium-tolerant rice varieties. **New Phytologist**, v. 154, n. 1, p. 131–145, 2002.

HALLBERG, Z. F.; SU, Y.; KITTO, R. Z.; HAMMOND, M. C. Engineering and *In Vivo* Applications of Riboswitches. **Annual Review of Biochemistry**, v. 86, n. 1, p. 515–539, 2017.

HANKINS, J. S.; DENROCHE, H.; MACKIE, G. A. Interactions of the RNA-binding protein Hfq with cspA mRNA, encoding the major cold shock protein. **Journal of bacteriology**, v. 192, n. 10, p. 2482–2490, 2010. Am Soc Microbiol.

HASSAN, S.; MATHESIUS, U. The role of flavonoids in root-rhizosphere signalling: Opportunities and challenges for improving plant-microbe interactions. **Journal of Experimental Botany**, v. 63, n. 9, p. 3429–3444, 2012.

HECKER, M.; LAMBECK, S.; TOEPFER, S.; VAN SOMEREN, E.; GUTHKE, R. Gene regulatory network inference: data integration in dynamic models—a review. **Biosystems**, v. 96, n. 1, p. 86–103, 2009.

HENKIN, T. M. Riboswitch RNAs: Using RNA to sense cellular metabolism. **Genes and Development**, v. 22, n. 24, p. 3383–3390, 2008.

HERBIG, A.; NIESELT, K. NocoRNAc: Characterization of non-coding RNAs in prokaryotes. **BMC Bioinformatics**, v. 12, 2011.

HERSHBERG, R.; ALTUVIA, S.; MARGALIT, H. A survey of small RNA-encoding genes in *Escherichia coli.* **Nucleic Acids Research**, v. 31, n. 7, p. 1813–1820, 2003.

HINDLEY, J. T. Fractionation of 32P-labelled ribonucleic acids on polyacrylamide gels and their characterization by fingerprinting. **Journal of molecular biology**, v. 30, n. 1, p. 125–136, 1967.

HOE, C.-H.; RAABE, C. A.; ROZHDESTVENSKY, T. S.; TANG, T.-H. Bacterial sRNAs: regulation in stress. **International Journal of Medical Microbiology**, v. 303, n. 5, p. 217–229, 2013.

HUYNH-THU, V. A.; IRRTHUM, A.; WEHENKEL, L.; GEURTS, P. Inferring regulatory networks from expression data using tree-based methods. **PLoS ONE**, v. 5, n. 9, p. 1–10, 2010.

JIANG, W.; FANG, L.; INOUYE, M. The role of the 5'-end untranslated region of the mRNA for CspA, the major cold-shock protein of *Escherichia coli*, in cold-shock adaptation. **Journal of Bacteriology**, v. 178, n. 16, p. 4919–4925, 1996.

JONES, B. D.; NICHOLS, W. A.; GIBSON, B. W.; SUNSHINE, M. G.; APICELLA, M. A. Study of the role of the htrB gene in *Salmonella typhimurium* virulence. **Infection and Immunity**, v. 65, n. 11, p. 4778–4783, 1997.

KÅHRSTRÖM, C. T. Bacterial Transcription: Introducing the excludon. **Nature Reviews Microbiology**, v. 10, n. 7, p. 443, 2012. Nature Publishing Group.

KALVARI, I.; ARGASINSKA, J.; QUINONES-OLVERA, N.; et al. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. **Nucleic acids research**, v. 46, n. D1, p. D335--D342, 2017. Oxford University Press.

KALVARI, I.; NAWROCKI, E. P.; ARGASINSKA, J.; et al. Non-Coding RNA Analysis Using the Rfam Database. **Current protocols in bioinformatics**, v. 62, n. 1, p. e51, 2018. Wiley Online Library.

KANDEL, S.; JOUBERT, P.; DOTY, S. Bacterial Endophyte Colonization and Distribution within Plants. **Microorganisms**, v. 5, n. 4, p. 77, 2017.

KANG, X.-M.; WANG, F.-F.; ZHANG, H.; ZHANG, Q.; QIAN, W. Genome-Wide Identification of Genes Necessary for Biofilm Formation by Nosocomial Pathogen *Stenotrophomonas maltophilia* Reveals that Orphan Response Regulator FsnR Is a Critical Modulator. **Applied and Environmental Microbiology**, v. 81, n. 4, p. 1200–1209, 2015.

KERY, M. B.; FELDMAN, M.; LIVNY, J.; TJADEN, B. TargetRNA2: Identifying targets of small regulatory RNAs in bacteria. **Nucleic Acids Research**, v. 42, n. W1, p. 124–129, 2014.

KHOO, J.; CHAI, S.; MOHAMED, R.; NATHAN, S.; FIRDAUS-RAIH, M. Computational discovery and RT-PCR validation of novel *Burkholderia* conserved and *Burkholderia pseudomallei* unique sRNAs. **BMC Genomics**, v. 13, n. Suppl 7, p. S13, 2012.

KIM, P. B.; NELSON, J. W.; BREAKER, R. R. An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. **Molecular cell**, v. 57, n. 2, p. 317–328, 2015.

KLASSEN, G.; PEDROSA, F. O.; SOUZA, E. M.; FUNAYAMA, S.; RIGO, L. U. Effect of nitrogen compounds on nitrogenase activity in *Herbaspirillum seropedicae* SmR1. **Canadian journal of microbiology**, v. 43, n. 9, p. 887–891, 1997.

KLEIN, R. J.; MISULOVIN, Z.; EDDY, S. R. Noncoding RNA genes identified in AT-rich hyperthermophiles. **Pnas**, v. 99, n. 11, p. 7542–7547, 2002.

DE LAY, N.; GOTTESMAN, S. A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. **Molecular Microbiology**, v. 86, n. 3, p. 524–538, 2012.

LENZ, D. H.; MOK, K. C.; LILLEY, B. N.; et al. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in Vibrio harveyi and Vibrio cholerae. **Cell**, v. 118, n. 1, p. 69–82, 2004.

LI, C.; LIU, S.; LIU, P.; et al. The YebN Leader RNA Acts as a Mn2+ Sensor Required for the Interaction of *Xanthomonas oryzae* and Rice. **Molecular plant-microbe interactions : MPMI**, p. 1–35, 2018.

LI, H.; HANDSAKER, B.; WYSOKER, A.; et al. The sequence alignment/map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 2009. Oxford University Press.

LIU, J. M.; CAMILLI, A. A broadening world of bacterial small RNAs. **Current opinion in microbiology**, v. 13, n. 1, p. 18–23, 2010.

LIU, X.; TIAN, J.; LIU, L.; et al. Identification of an operon involved in fluoride resistance in *Enterobacter cloacae* FRM. **Scientific Reports**, v. 7, n. 1, p. 1–9, 2017. Springer US.

LOPES, F. M.; CESAR JR, R. M.; COSTA, L. D. F. Gene expression complex networks: synthesis, identification, and analysis. **Journal of Computational Biology**, v. 18, n. 10, p. 1353–1367, 2011. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA.

LOPES, F. M.; MARTINS-JR, D. C.; CESAR-JR, R. M. Comparative study of GRN'S inference methods based on feature selection by mutual information. **2009 IEEE International Workshop on Genomic Signal Processing and Statistics, GENSIPS 2009**, p. 1–4, 2009. IEEE.

LOPES, F. M.; MARTINS, D. C.; CESAR, R. M. Feature selection environment for genomic applications. **BMC bioinformatics**, v. 9, n. 1, p. 451, 2008. BioMed Central.

LOPES, F. M.; MARTINS JR, D. C.; BARRERA, J.; CESAR JR, R. M. A feature selection technique for inference of graphs from their known topological properties: Revealing scale-free gene regulatory networks. **Information Sciences**, v. 272, p. 1–15, 2014.

LOPES, F. M.; MENDONÇA, V. R.; ANDRADE, B. B.; BARRAL-NETTO, M.; QUEIROZ, A. T. Networking the host immune response in *Plasmodium vivax* malaria. **Malaria Journal**, v. 12, n. 1, p. 69, 2013.

LU, X.; GOODRICH-BLAIR, H.; TJADEN, B. Assessing computational tools for the discovery of small RNA genes in bacteria. **RNA**, v. 17, n. 9, p. 1635–1647, 2011. Cold Spring

Harbor Lab.

MAJDALANI, N.; VANDERPOOL, C. K.; GOTTESMAN, S. Bacterial small RNA regulators. **Critical Reviews in Biochemistry and Molecular Biology**, v. 40, n. 2, p. 93–113, 2005.

MANN, M.; WRIGHT, P. R.; BACKOFEN, R. IntaRNA 2.0: Enhanced and customizable prediction of RNA-RNA interactions. **Nucleic Acids Research**, v. 45, n. W1, p. W435–W439, 2017.

MARBACH, D.; COSTELLO, J. C.; KÜFFNER, R.; et al. Wisdom of crowds for robust gene network inference. **Nature Methods**, v. 9, n. 8, p. 796–804, 2012.

MARGOLIN, A. A.; NEMENMAN, I.; BASSO, K.; et al. ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. **BMC Bioinformatics**, v. 7, n. SUPPL.1, p. 1–15, 2006.

MARGOLIS, P. S.; HACKBARTH, C. J.; YOUNG, D. C.; et al. Peptide deformylase in *Staphylococcus aureus*: resistance to inhibition is mediated by mutations in the formyltransferase gene. **Antimicrobial agents and chemotherapy**, v. 44, n. 7, p. 1825–1831, 2000. Am Soc Microbiol.

MARGUERAT, S.; BÄHLER, J. RNA-seq: from technology to biology. **Cellular and molecular life sciences**, v. 67, n. 4, p. 569–579, 2010. Springer.

MARIN, A. M.; SOUZA, E. M.; PEDROSA, F. O.; et al. Naringenin degradation by the endophytic diazotroph *Herbaspirillum seropedicae* SmR1. **Microbiology**, v. 159, n. 1, p. 167–175, 2013. Microbiology Society.

MASSÉ, E.; ESCORCIA, F. E.; GOTTESMAN, S. Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. **Genes and Development**, v. 17, n. 19, p. 2374–2383, 2003.

MATOS, R. G.; CASINHAS, J.; BÁRRIA, C.; et al. The role of ribonucleases and sRNAs in the virulence of foodborne pathogens. **Frontiers in Microbiology**, v. 8, n. MAY, p. 1–7, 2017.

MEHDIZADEH AGHDAM, E.; SINN, M.; TARHRIZ, V.; et al. TPP riboswitch characterization in *Alishewanella tabrizica* and *Alishewanella aestuarii* and comparison with other TPP riboswitches. **Microbiological Research**, v. 195, p. 71–80, 2017.

MEISIG, J.; BLÜTHGEN, N. The gene regulatory network of mESC differentiation: A benchmark for reverse engineering methods. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 373, n. 1750, 2018.

MELAMED, S.; PEER, A.; FAIGENBAUM-ROMM, R.; et al. Global Mapping of Small RNA-Target Interactions in Bacteria. **Molecular Cell**, v. 63, n. 5, p. 884–897, 2016.

MELLIN, J. R.; COSSART, P. The non-coding RNA world of the bacterial pathogen *Listeria monocytogenes*. **RNA biology**, v. 9, n. 4, p. 372–378, 2012. Taylor & Francis.

MELLIN, J. R.; COSSART, P. Unexpected versatility in bacterial riboswitches. **Trends in Genetics**, v. 31, n. 3, p. 150–156, 2015.

MERCIER, E.; HOLTKAMP, W.; RODNINA, M. V.; WINTERMEYER, W. Signal recognition particle binds to translating ribosomes before emergence of a signal anchor sequence. **Nucleic acids research**, v. 45, n. 20, p. 11858–11866, 2017. Oxford University Press.

MEYER, P. E.; LAFITTE, F.; BONTEMPI, G. Minet: A r/bioconductor package for inferring large transcriptional networks using mutual information. **BMC Bioinformatics**, v. 9, p. 1–10, 2008.

MICHAUX, C.; VERNEUIL, N.; HARTKE, A.; GIARD, J. C. Physiological roles of small RNA molecules. **Microbiology (United Kingdom)**, v. 160, n. PART 6, p. 1007–1019, 2014.

MIKA, F.; HENGGE, R. Small RNAs in the control of RpoS, CsgD, and biofilm architecture of *Escherichia coli*. **RNA Biology**, v. 11, n. 5, 2014.

MINAMINO, T.; KINOSHITA, M.; NAMBA, K. Fuel of the bacterial flagellar type III protein export apparatus. **The Bacterial Flagellum**. p.3–16, 2017. Springer.

MINAMINO, T.; MACNAB, R. M. FliH, a soluble component of the type III flagellar export apparatus of *Salmonella*, forms a complex with FliI and inhibits its ATPase activity. **Molecular microbiology**, v. 37, n. 6, p. 1494–1503, 2000. Wiley Online Library.

MODI, S. R.; CAMACHO, D. M.; KOHANSKI, M. A.; WALKER, G. C.; COLLINS, J. J. Functional characterization of bacterial sRNAs using a network biology approach. **Proceedings of the National Academy of Sciences**, v. 108, n. 37, p. 15522–15527, 2011.

MONTEIRO, R. A.; SOUZA, E. M.; FUNAYAMA, S.; et al. Expression and functional analysis of an N-truncated NifA protein of *Herbaspirillum seropedicae*. **FEBS letters**, v. 447, n. 2–3, p. 283–286, 1999. Wiley Online Library.

MORENO, L. F. Predição e análise da expressão de RNAS não codificadores com função regulatória presentes na bactéria *Herbaspirillum seropedicae* SmR1. , 2013.

NARBERHAUS, F.; WALDMINGHAUS, T.; CHOWDHURY, S. RNA thermometers. **FEMS Microbiology Reviews**, v. 30, n. 1, p. 3–16, 2006.

NAWAZ, M. Z.; JIAN, H.; HE, Y.; et al. Genome-wide detection of small regulatory RNAs in deep-sea bacterium *Shewanella piezotolerans* WP3. **Frontiers in Microbiology**, v. 8, p. 1093, 2017. Frontiers.

NAWROCKI, E. P.; EDDY, S. R. Infernal 1.1: 100-fold faster RNA homology searches. **Bioinformatics**, v. 29, n. 22, p. 2933–2935, 2013. Oxford University Press.

NELSON, J. W.; ATILHO, R. M.; SHERLOCK, M. E.; STOCKBRIDGE, R. B.; BREAKER, R. R. Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. **Molecular cell**, v. 65, n. 2, p. 220–230, 2017.

NITZAN, M.; REHANI, R.; MARGALIT, H. Integration of Bacterial Small RNAs in Regulatory Networks. **Annual Review of Biophysics**, v. 46, n. 1, p. 131–148, 2017.

OLIVA, G.; SAHR, T.; BUCHRIESER, C. Small RNAs, 5' UTR elements and RNA-binding proteins in intracellular bacteria: Impact on metabolism and virulence. **FEMS Microbiology Reviews**, v. 39, n. 3, p. 331–349, 2015.

OLIVARES, F. L.; BALDANI, V. L. D.; REIS, V. M.; BALDANI, J. I.; DÖBEREINER, J. Occurrence of the endophytic diazotrophs *Herbaspirillum* spp. in roots, stems, and leaves, predominantly of Gramineae. **Biology and Fertility of Soils**, v. 21, n. 3, p. 197–200, 1996. Springer.

PAIN, A.; OTT, A.; AMINE, H.; et al. An assessment of bacterial small RNA target prediction programs. **RNA Biology**, v. 12, n. 5, p. 509–513, 2015.

PANKIEVICZ, V. C. S.; CAMILIOS-NETO, D.; BONATO, P.; et al. RNA-seq transcriptional profiling of *Herbaspirillum seropedicae* colonizing wheat (*Triticum aestivum*) roots. **Plant Molecular Biology**, v. 90, n. 6, p. 589–603, 2016. Springer Netherlands.

PAPENFORT, K.; FÖRSTNER, K. U.; CONG, J.-P.; SHARMA, C. M.; BASSLER, B. L. Differential RNA-seq of Vibrio cholerae identifies the VqmR small RNA as a regulator of biofilm formation. **Proceedings of the National Academy of Sciences**, v. 112, n. 7, p. E766--E775, 2015. National Acad Sciences.

PAUL SHANNON, 1; ANDREW MARKIEL, 1; OWEN OZIER, 2 NITIN S. BALIGA, 1 JONATHAN T. WANG, 2 DANIEL RAMAGE, 2; et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. **Genome Research**, v. 13, n. 22, p. 6, 2003.

PEDROLLI, D.; LANGER, S.; HOBL, B.; et al. The ribB FMN riboswitch from *Escherichia coli* operates at the transcriptional and translational level and regulates riboflavin biosynthesis. **FEBS Journal**, v. 282, n. 16, p. 3230–3242, 2015.

PEDROSA, F. O.; MONTEIRO, R. A.; WASSEM, R.; et al. Genome of *Herbaspirillum seropedicae* strain SmR1, a specialized diazotrophic endophyte of tropical grasses. **PLoS Genetics**, v. 7, n. 5, 2011.

PHONGSISAY, V.; PERERA, V. N.; FRY, B. N. Expression of the htrB gene is essential for responsiveness of *Salmonella typhimurium* and *Campylobacter jejuni* to harsh environments. **Microbiology**, v. 153, n. 1, p. 254–262, 2007.

PIMENTEL, J. P.; OLIVARES, F.; PITARD, R. M.; et al. Dinitrogen fixation and infection of grass leaves by *Pseudomonas rubrisubalbicans* and *Herbaspirillum seropedicae*. **Nitrogen Fixation**. p.225–229, 1991. Springer.

POLASKI, J. T.; WEBSTER, S. M.; JOHNSON, J. E.; BATEY, R. T. Cobalamin riboswitches exhibit a broad range of ability to discriminate between methylcobalamin and adenosylcobalamin. **Journal of Biological Chemistry**, v. 292, n. 28, p. 11650–11658, 2017.

PRICE, I. R.; GABALLA, A.; DING, F.; HELMANN, J. D.; KE, A. Mn2+-sensing mechanisms of yybP-ykoY orphan riboswitches. **Molecular cell**, v. 57, n. 6, p. 1110–1123, 2015.

RANJAN, A.; MERCIER, E.; BHATT, A.; WINTERMEYER, W. Signal recognition particle prevents N-terminal processing of bacterial membrane proteins. **Nature Communications**, v. 8, n. May, p. 1–6, 2017.

LE RHUN, A.; BEER, Y. Y.; REIMEGÅRD, J.; CHYLINSKI, K.; CHARPENTIER, E. RNA sequencing uncovers antisense RNAs and novel small RNAs in Streptococcus pyogenes. **RNA biology**, v. 13, n. 2, p. 177–195, 2016. Taylor & Francis.

RICHARDS, G. R.; VANDERPOOL, C. K. Molecular call and response: The physiology of bacterial small RNAs. **Biochimica et Biophysica Acta - Gene Regulatory Mechanisms**, v. 1809, n. 10, p. 525–531, 2011.

RIVAS, E.; EDDY, S. R. Noncoding RNA gene detection using comparative sequence analysis. **BMC bioinformatics**, v. 2, n. 1, p. 8, 2001. BioMed Central.

RODIONOV, D. A.; VITRESCHAK, A. G.; MIRONOV, A. A.; GELFAND, M. S. Comparative Genomics of the Vitamin B12 Metabolism and Regulation in Prokaryotes. **Journal of Biological Chemistry**, v. 278, n. 42, p. 41148–41159, 2003.

RONCATO-MACCARI, L. D. B.; RAMOS, H. J. O.; PEDROSA, F. O.; et al. Endophytic *Herbaspirillum seropedicae* expresses nif genes in gramineous plants. **FEMS Microbiology Ecology**, v. 45, n. 1, p. 39–47, 2003.

RUMBLE, S. M.; LACROUTE, P.; DALCA, A. V; et al. SHRiMP: accurate mapping of short color-space reads. **PLoS computational biology**, v. 5, n. 5, p. e1000386, 2009. Public Library of Science.

SABERI, F.; KAMALI, M.; NAJAFI, A.; YAZDANPARAST, A.; MOGHADDAM, M. M. Natural antisense RNAs as mRNA regulatory elements in bacteria: a review on function and applications. **Cellular & Molecular Biology Letters**, v. 21, n. 1, p. 6, 2016.

SAJÓ, R.; LILIOM, K.; MUSKOTÁL, A.; et al. Soluble components of the flagellar export apparatus, FliI, FliJ, and FliH, do not deliver flagellin, the major filament protein, from the cytosol to the export gate. **Biochimica et Biophysica Acta - Molecular Cell Research**, v. 1843, n. 11, p. 2414–2423, 2014.

SARAMAGO, M.; BÁRRIA, C.; DOS SANTOS, R. F.; et al. The role of RNases in the regulation of small RNAs. **Current Opinion in Microbiology**, v. 18, n. 1, p. 105–115, 2014.

SCHMIDT, M. A.; SOUZA, E. M.; BAURA, V.; et al. Evidence for the endophytic colonization of *Phaseolus vulgaris* (common bean) roots by the diazotroph *Herbaspirillum seropedicae*. **Brazilian Journal of Medical and Biological Research**, v. 44, n. 3, p. 182–185, 2011.

SCHULTZE, T.; HILKER, R.; MANNALA, G. K.; et al. A detailed view of the intracellular transcriptome of *Listeria monocytogenes* in murine macrophages using RNA-seq. **Frontiers in microbiology**, v. 6, p. 1199, 2015. Frontiers.

SERGANOV, A.; NUDLER, E. A decade of riboswitches. **Cell**, v. 152, n. 1–2, p. 17–24, 2013.

SHARMA, A.; KHULLER, G. K.; SHARMA, S. Peptide deformylase--a promising therapeutic target for tuberculosis and antibacterial drug discovery. **Expert opinion on therapeutic targets**, v. 13, n. 7, p. 753–765, 2009. Taylor & Francis.

SHARMA, C. M.; HEIDRICH, N. Small RNAs and virulence in bacterial pathogens. **RNA Biology**, v. 9, n. 4, p. 361–363, 2012.

SILVA, I. J.; BARAHONA, S.; EYRAUD, A.; et al. SraL sRNA interaction regulates the terminator by preventing premature transcription termination of rho mRNA . **Proceedings of the National Academy of Sciences**, v. 116, n. 8, p. 3042–3051, 2019.

SKIPPINGTON, E.; RAGAN, M. A. Evolutionary dynamics of small RNAs in 27 *Escherichia coli* and *Shigella genomes*. **Genome biology and evolution**, v. 4, n. 3, p. 330–345, 2012. Oxford University Press.

SOUZA, E. M.; PEDROSA, F. O.; RIGO, L. U.; MACHADO, H. B.; YATES, M. G. Expression of the nifA gene of *Herbaspirillum seropedicae*: role of the NtrC and NifA binding sites and of the- 24/- 12 promoter element. **Microbiology**, v. 146, n. 6, p. 1407–1418, 2000. Microbiology Society.

SRIDHAR, J.; GUNASEKARAN, P. Computational small RNA prediction in bacteria. **Bioinformatics and Biology Insights**, v. 7, p. 83–95, 2013.

STEUTEN, B.; HOCH, P. G.; DAMM, K.; et al. Regulation of transcription by 6S RNAs insights from the *Escherichia coli* and *Bacillus subtilis* model systems. **RNA Biology**, v. 11, n. 5, 2014.

STORZ, G.; VOGEL, J.; WASSARMAN, K. M. Regulation by small RNAs in bacteria: expanding frontiers. **Molecular cell**, v. 43, n. 6, p. 880–891, 2011.

SUBRAMANIAN, D.; BHASURAN, B.; NATARAJAN, J. Genomic analysis of RNA-Seq and sRNA-Seq data identifies potential regulatory sRNAs and their functional roles in *Staphylococcus aureus*. **Genomics**, , n. September, p. 1–16, 2018.

SUN, E. I.; LEYN, S. A.; KAZANOV, M. D.; et al. Comparative genomics of metabolic capacities of regulons controlled by cis-regulatory RNA motifs in bacteria. **BMC genomics**, v. 14, n. 1, p. 597, 2013. BioMed Central.

SUNDERMEIER, T. R.; DULEBOHN, D. P.; CHO, H. J.; KARZAI, A W. A previously uncharacterized role for small protein B (SmpB) in transfer messenger RNA-mediated trans-translation. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 7, p. 2316–21, 2005.

TADRA-SFEIR, M. Z.; FAORO, H.; CAMILIOS-NETO, D.; et al. Genome wide transcriptional profiling of Herbaspirillum seropedicae SmR1 grown in the presence of naringenin. **Frontiers in Microbiology**, v. 6, n. MAY, p. 1–8, 2015.

TATUSOV, R. L.; GALPERIN, M. Y.; NATALE, D. A.; KOONIN, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. **Nucleic acids research**, v. 28, n. 1, p. 33–6, 2000.

TRAPNELL, C.; ROBERTS, A.; GOFF, L.; et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nature Protocols**, v. 7, n. 3, p. 562–578, 2012.

TROTOCHAUD, A. E.; WASSARMAN, K. M. A highly conserved 6S RNA structure is required for regulation of transcription. **Nature Structural and Molecular Biology**, v. 12, n. 4, p. 313–319, 2005.

VOGEL, J.; SHARMA, C. M. How to find small non-coding RNAs in bacteria. **Biological Chemistry**, v. 386, n. 12, p. 1219–1238, 2005.

VOGEL, J.; WAGNER, E. G. H. Target identification of small noncoding RNAs in bacteria. **Current Opinion in Microbiology**, v. 10, n. 3, p. 262–270, 2007.

WANG, J. X.; LEE, E. R.; MORALES, D. R.; LIM, J.; BREAKER, R. R. Riboswitches that Sense S-adenosylhomocysteine and Activate Genes Involved in Coenzyme Recycling. **Molecular Cell**, v. 29, n. 6, p. 691–702, 2008.

WASSARMAN, K. M.; REPOILA, F.; ROSENOW, C.; STORZ, G.; GOTTESMAN, S. Identification of novel small RNAs using comparative genomics and microarrays. **Genes and Development**, v. 15, n. 13, p. 1637–1651, 2001.

WASSARMAN, K. M.; STORZ, G. 6S RNA regulates *E. coli* RNA polymerase activity. **Cell**, v. 101, n. 6, p. 613–623, 2000.

WATERS, L. S.; STORZ, G. Regulatory RNAs in Bacteria. **Cell**, v. 136, n. 4, p. 615–628, 2009.

WEISS, V. A. Estratégias de finalização da montagem do genoma da Bactéria Diazotrófica Endofítica *Herbaspirillum seropedicae* SmR1, 2010.

WINKLER, W. C.; BREAKER, R. R. Regulation of Bacterial Gene Expression By Riboswitches. **Annual Review of Microbiology**, v. 59, n. 1, p. 487–517, 2005.

WONG, J.; PANG, I.; WILKINS, M.; TREE, J. J. **Systems-Level Analysis of Bacterial Regulatory Small RNA Networks**. 2018.

ZARATIEGUI, M.; IRVINE, D. V.; MARTIENSSEN, R. A. Noncoding RNAs and Gene Silencing. **Cell**, v. 128, n. 4, p. 763–776, 2007.

ZHAN, S.; DONG, Y.; ZHAO, W.; et al. Genome-wide identification and characterization of long non-coding RNAs in developmental skeletal muscle of fetal goat. **BMC Genomics**, v. 17, n. 1, p. 1–10, 2016.

ZHANG, J.; LU, K.; XIANG, Y.; et al. Weighted frequent gene co-expression network mining to identify genes involved in genome stability. **PLoS computational biology**, v. 8, n. 8, p. e1002656, 2012. Public Library of Science.

ZHANG, Y.; BURKHARDT, D. H.; ROUSKIN, S.; et al. A Stress Response that Monitors and Regulates mRNA Structure Is Central to Cold Shock Adaptation. **Molecular Cell**, v. 70, n. 2, p. 274- 286.e7, 2018.

ZHAO, T.; ZHANG, R.; WANG, M. Prediction of candidate small non-coding RNAs in by *Agrobacterium* by computational analysis. **Journal of biomedical research**, v. 24, n. 1, p. 33–42, 2010.

ZHU, P.; HAN, J. Asynchronous stochastic Boolean networks as gene network models. **Journal of Computational Biology**, v. 21, n. 10, p. 771–783, 2014. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA.

ZORGANI, M. A.; QUENTIN, R.; LARTIGUE, M. F. Regulatory RNAs in the less studied *Streptococcal* species: From nomenclature to identification. **Frontiers in Microbiology**, v. 7, n. JUL, 2016.